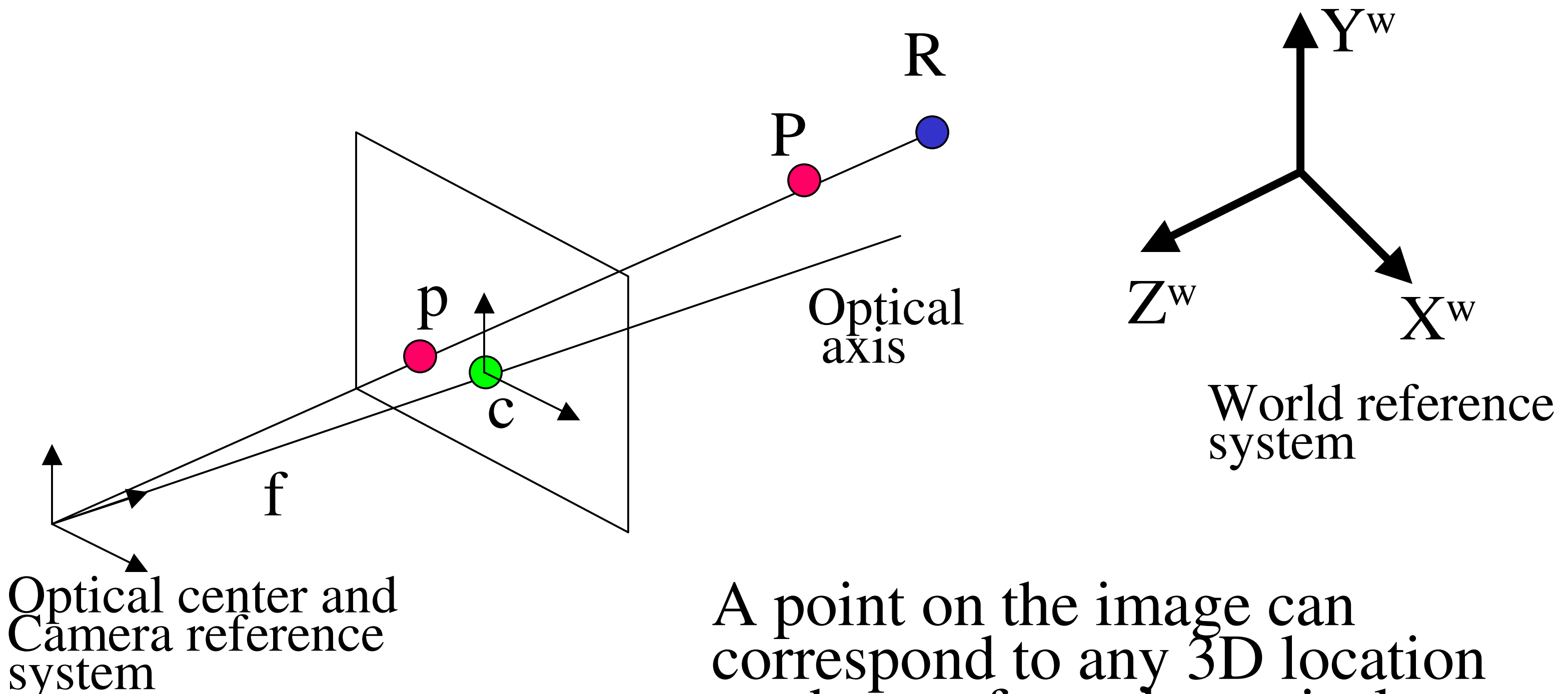
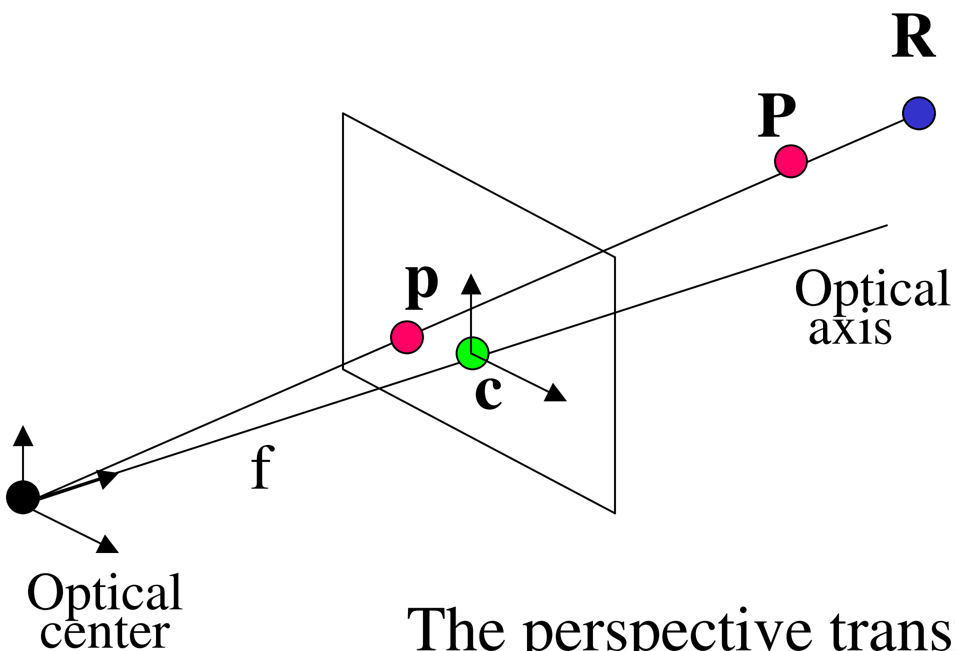
Stereo Vision

The inverse perspective (1)



A point on the image can correspond to any 3D location on the ray from the optical center through that point on the image plane.

The inverse perspective (2)



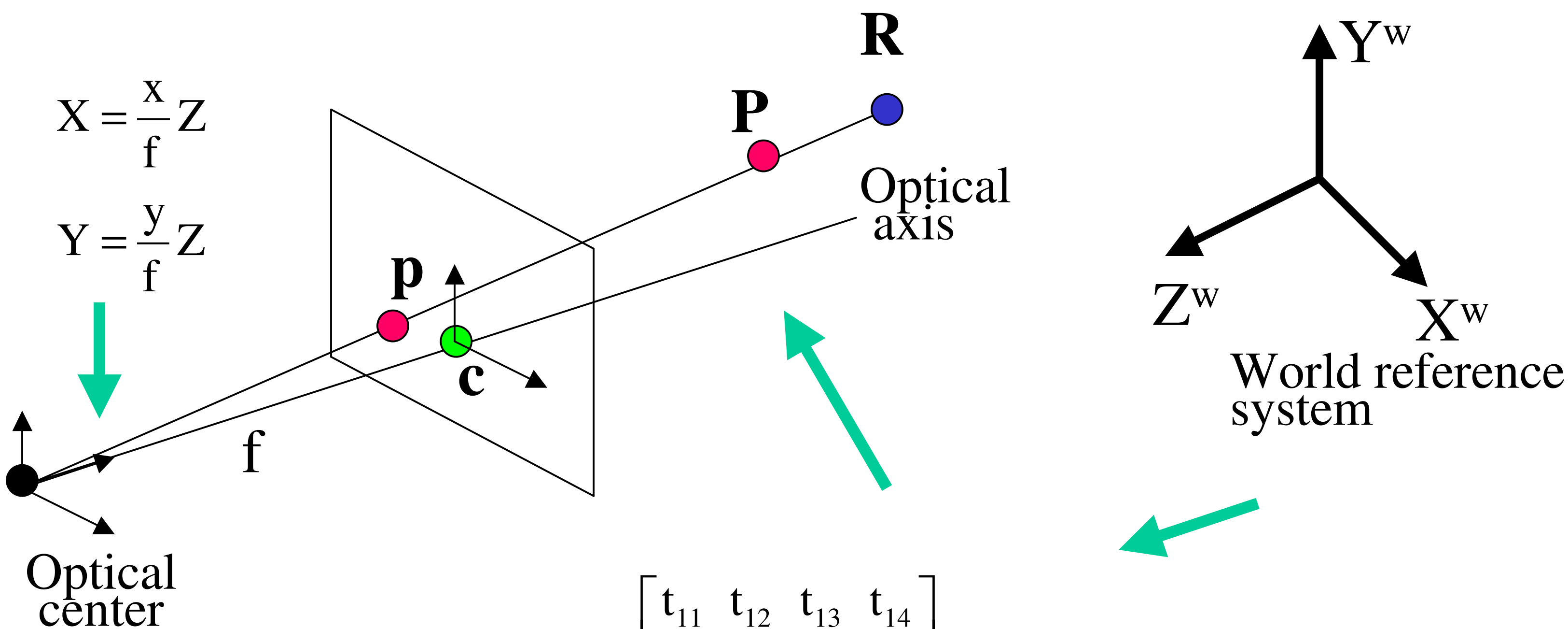
$(x, y, 0) \Rightarrow$ image coord.

$$X = \frac{x}{f} Z$$

$$Y = \frac{x}{f} Z$$

The perspective transformation is not invertible, however, it is possible to calculate a vector pointing in the direction of the ray from the optical center through that point on the image plane

The inverse perspective reference system



$$\begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix}$$

camera projection matrix

The inverse perspective in World Reference System

$$\begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix}$$

camera projection matrix

$(x, y) \Rightarrow$ image coordinates

$$a_1 = t_{11} - xt_{31} \qquad a_2 = t_{21} - yt_{31}$$
$$b_1 = t_{12} - xt_{32} \qquad b_2 = t_{22} - yt_{32}$$
$$c_1 = t_{13} - xt_{33} \qquad c_2 = t_{23} - yt_{33}$$
$$d_1 = t_{14} - xt_{34} \qquad d_2 = t_{24} - yt_{34}$$

$$\frac{X - x_0}{\lambda} = \frac{Y - y_0}{\mu} = \frac{Z - z_0}{\rho}$$

Parametric line in WRS

$$(\lambda, \mu, \rho) = (b_1c_2 - b_2c_1, c_1a_2 - c_2a_1, a_1b_2 - a_2b_1)$$

$$x_0 = \frac{b_1(c_2z_0 + d_2) - b_2(c_1z_0 + d_1)}{a_1b_2 - b_1a_2}$$

$$y_0 = \frac{a_2(c_1z_0 + d_1) - a_1(c_2z_0 + d_2)}{a_1b_2 - b_1a_2}$$

The inverse perspective in World Reference System (1)

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix}$$

camera projection matrix

$(x, y) \Rightarrow$ image coordinates

$P(X, Y, Z) \Rightarrow$ point in WRS

$$t_1 = \begin{bmatrix} t_{i1} & t_{i2} & t_{i3} \end{bmatrix}^T$$

t_i vector composed of the first three elements of the row i of T

An image point I is associated with a scene point P by the equation

$$I = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = T\,P$$

$$\mathbf{P^T t_1} + t_{14} - x(\mathbf{P^T t_3} + 1) = \mathbf{0}$$

$$\mathbf{P^T t_2} + t_{24} - y(\mathbf{P^T t_3} + 1) = \mathbf{0}$$

$$\Downarrow$$

$$(\mathbf{t_1} - x\,\mathbf{t_3})^T \mathbf{P} + (t_{14} - x\,t_{34}) = 0$$

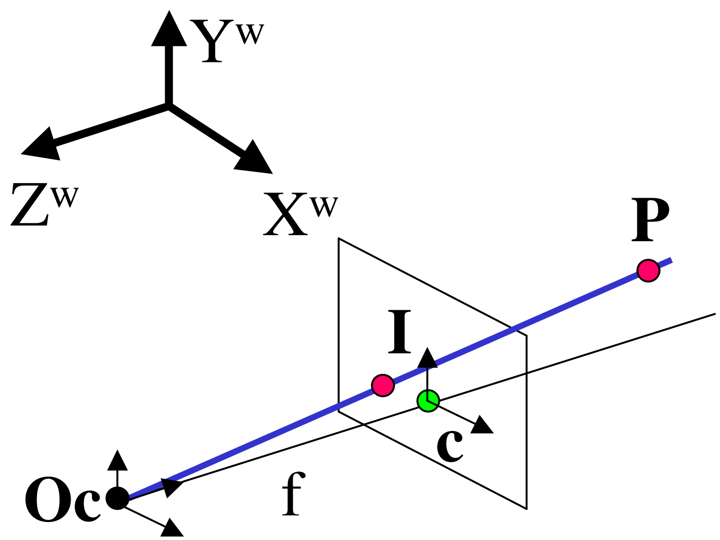
$$(\mathbf{t_2} - y\,\mathbf{t_3})^T \mathbf{P} + (t_{24} - y\,t_{34}) = 0$$

The inverse perspective in World Reference System (2)

$$\left.\begin{aligned}(\mathbf{t_1} - x\,\mathbf{t_3})^T\,\mathbf{P} + (t_{14} - x\,t_{34}) = 0\\(\mathbf{t_2} - y\,\mathbf{t_3})^T\,\mathbf{P} + (t_{24} - y\,t_{34}) = 0\end{aligned}\right\}$$

Equations of two planes whose intersection is the line OcI (expressed in the WRS)

The direction vector n of this line is the vector product of normal vector of the two planes



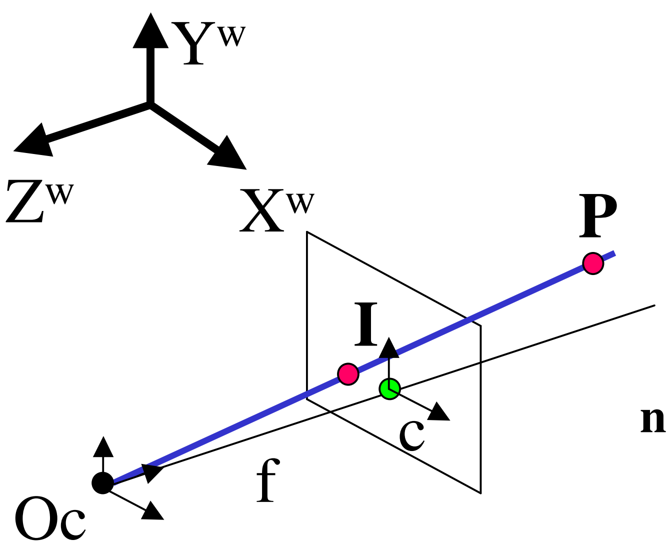
$$\mathbf{n} = (\mathbf{t_1} - \mathrm{x}\,\mathbf{t_3}) \times (\mathbf{t_2} - \mathrm{y}\,\mathbf{t_3})$$

i.e.,

$$\mathbf{n} = \mathrm{x}(\mathbf{t_2} \times \mathbf{t_3}) + \mathrm{y}(\mathbf{t_3} \times \mathbf{t_1}) + (\mathbf{t_1} \times \mathbf{t_2})$$

$$\mathbf{n} = \left[(\mathbf{t_2} \times \mathbf{t_3})\ \ (\mathbf{t_3} \times \mathbf{t_1})\ \ (\mathbf{t_1} \times \mathbf{t_2})\right]\begin{bmatrix}\mathrm{x}\\ \mathrm{y}\\ 1\end{bmatrix}$$

The inverse perspective in World Reference System (3)



$$\mathbf{n} = \left[(\mathbf{t_2} \times \mathbf{t_3})\ \ (\mathbf{t_3} \times \mathbf{t_1})\ \ (\mathbf{t_1} \times \mathbf{t_2})\right] \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = S \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

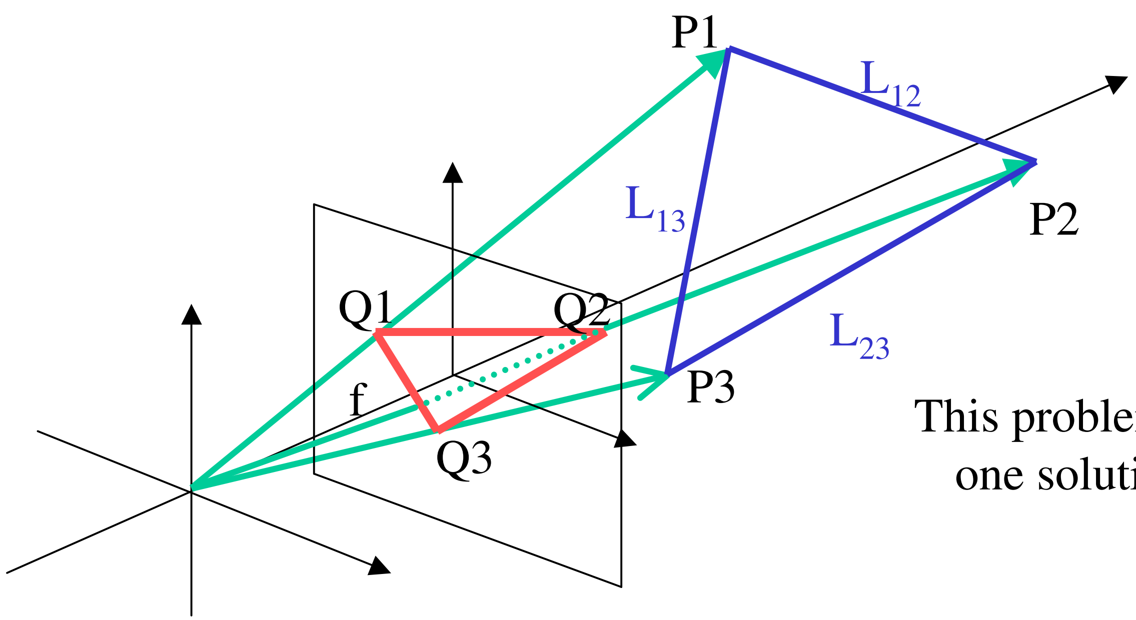
$$S = \begin{bmatrix} (t_{22}t_{33} - t_{32}t_{23}) & (t_{32}t_{13} - t_{12}t_{33}) & (t_{12}t_{23} - t_{22}t_{13}) \\ (t_{31}t_{23} - t_{21}t_{33}) & (t_{11}t_{33} - t_{31}t_{13}) & (t_{21}t_{13} - t_{11}t_{23}) \\ (t_{21}t_{32} - t_{31}t_{22}) & (t_{31}t_{12} - t_{11}t_{32}) & (t_{11}t_{22} - t_{21}t_{12}) \end{bmatrix}$$

$$\mathbf{n} = \begin{bmatrix} x(t_{22}t_{33} - t_{32}t_{23}) + y(t_{32}t_{13} - t_{12}t_{33}) + (t_{12}t_{23} - t_{22}t_{13}) \\ x(t_{31}t_{23} - t_{21}t_{33}) + y(t_{11}t_{33} - t_{31}t_{13}) + (t_{21}t_{13} - t_{11}t_{23}) \\ x(t_{21}t_{32} - t_{31}t_{22}) + y(t_{31}t_{12} - t_{11}t_{32}) + (t_{11}t_{22} - t_{21}t_{12}) \end{bmatrix}$$

The parametric equation of the line OcI in WRS is: $\mathbf{P} = \mathbf{Oc} + \lambda \mathbf{n}$

Depth information from single image

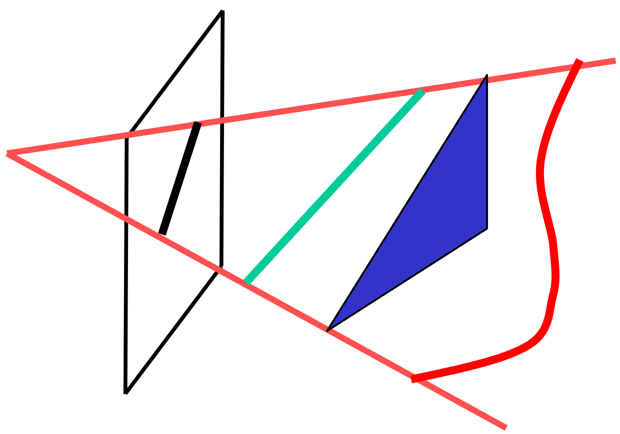
- 2D imaging results in depth information loss
- If only one camera is used, the 3-D depth information can be derived if prior knowledge is available



This problem has more than one solution: ambiguity

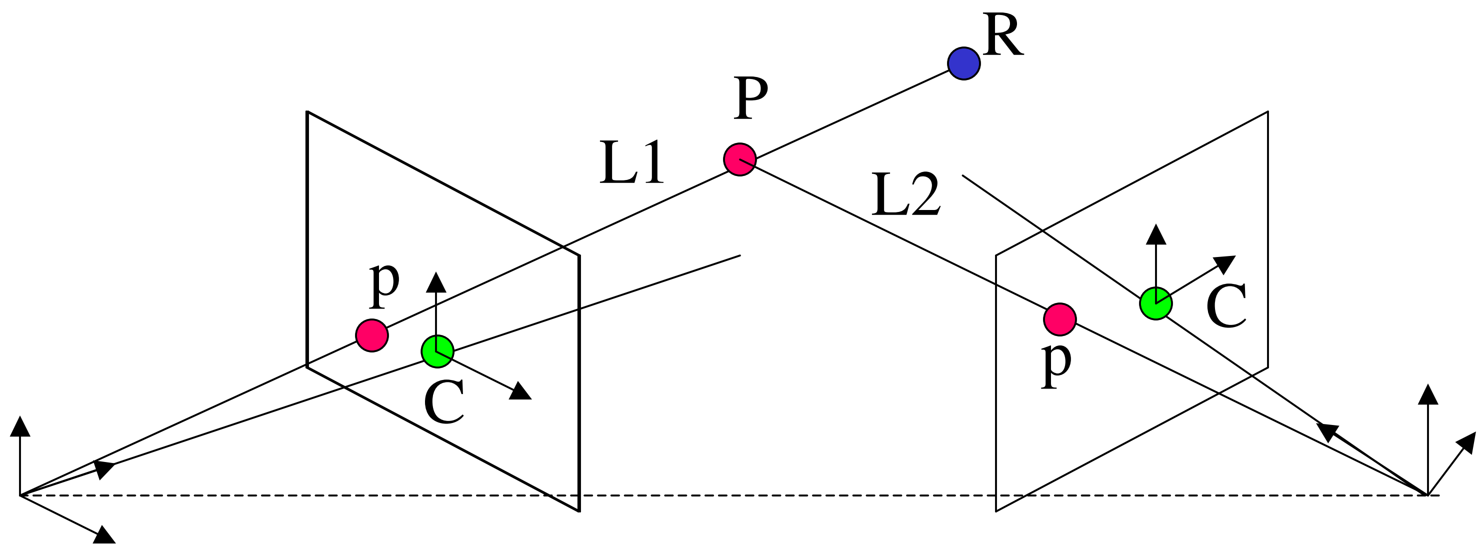
Single view ambiguity

- Single views are insufficient to solve geometric problems computer vision.
- It is necessary methods of understanding multiple views



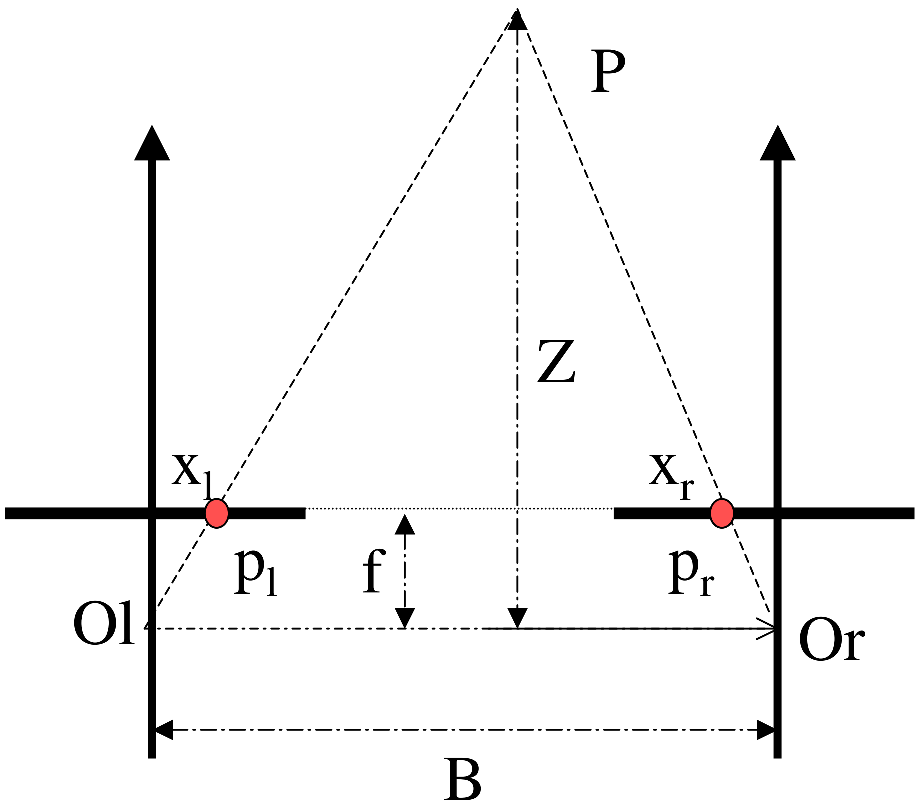
- **Stereo**: different cameras, different viewpoints, same time
- **Motion**: same camera, different viewpoints, different times

Depth information from a pair of images



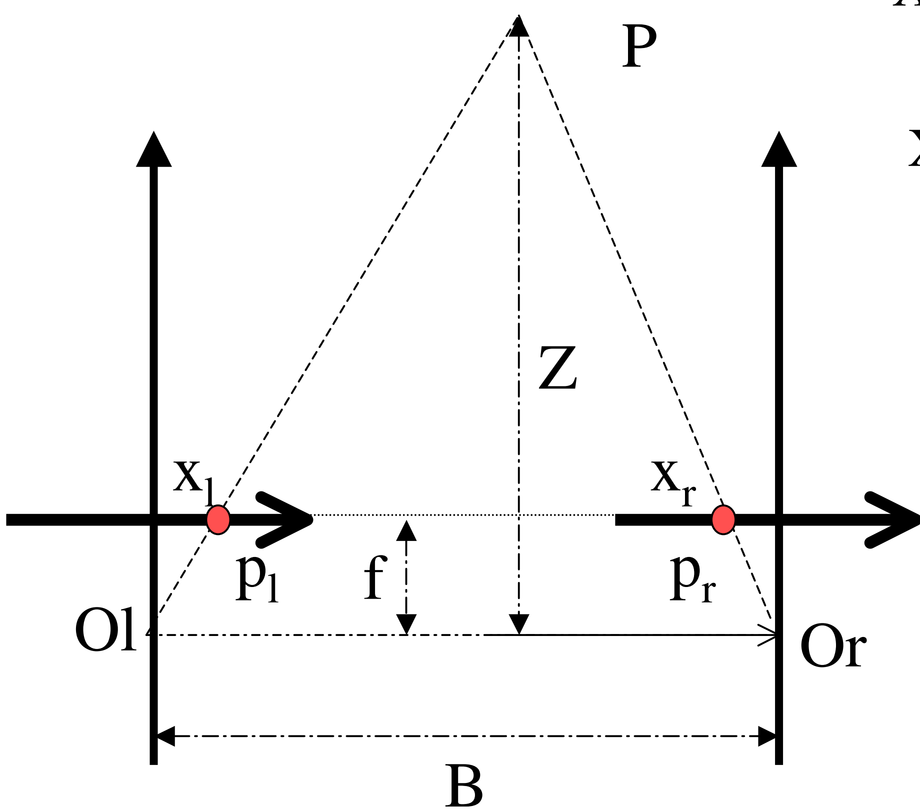
- Lines L1 and L2: obtained by inverse perspective
- Point P: determined by the intersection of L1 and L2 (triangulation)

A simple stereo system (1)



- Image planes of cameras are parallel
- Focal points are at same height
- Focal lengths same

A simple stereo system (2)



$$X_l = \frac{x_l}{f} Z_l \qquad (1) \qquad\qquad X_r = \frac{x_r}{f} Z_r \qquad (2)$$

$$X_r = X_r + B \qquad (3) \qquad Z = Z_l = Z_r \qquad (4)$$

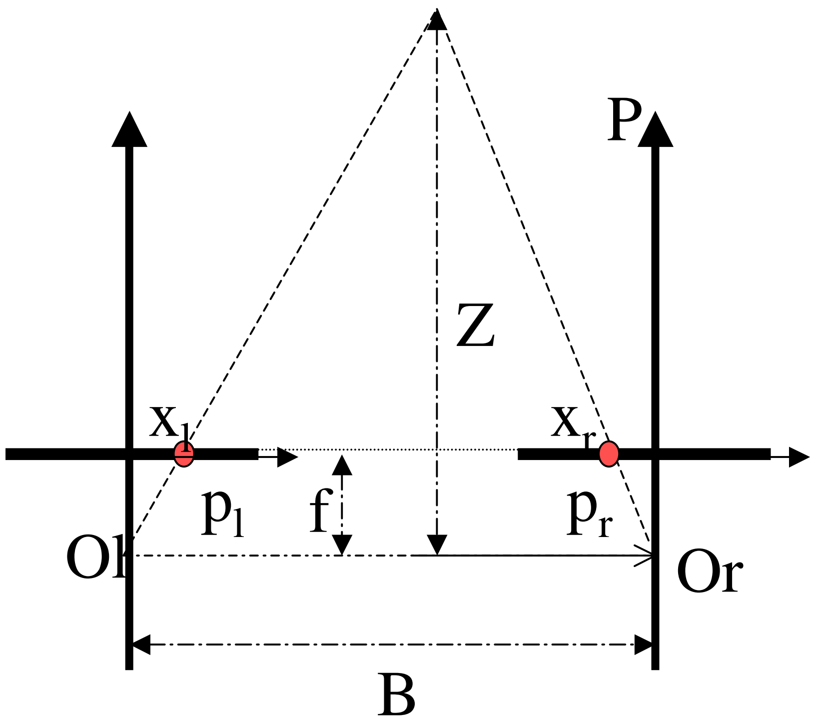
$$(4) \rightarrow (1) \quad X_l = \frac{x_l}{f} Z \qquad (5)$$

$$(3) \rightarrow (5) \quad X_r + B = \frac{x_l}{f} Z \qquad (6)$$

$$(2)(4) \rightarrow (6) \quad \frac{x_r}{f} Z + B = \frac{x_l}{f} Z \qquad (7)$$

$$Z = \frac{f\,B}{x_l - x_r} \qquad (8)$$

A simple stereo system (3)



$$Z = \frac{f\,B}{x_l - x_r} \quad (8)$$

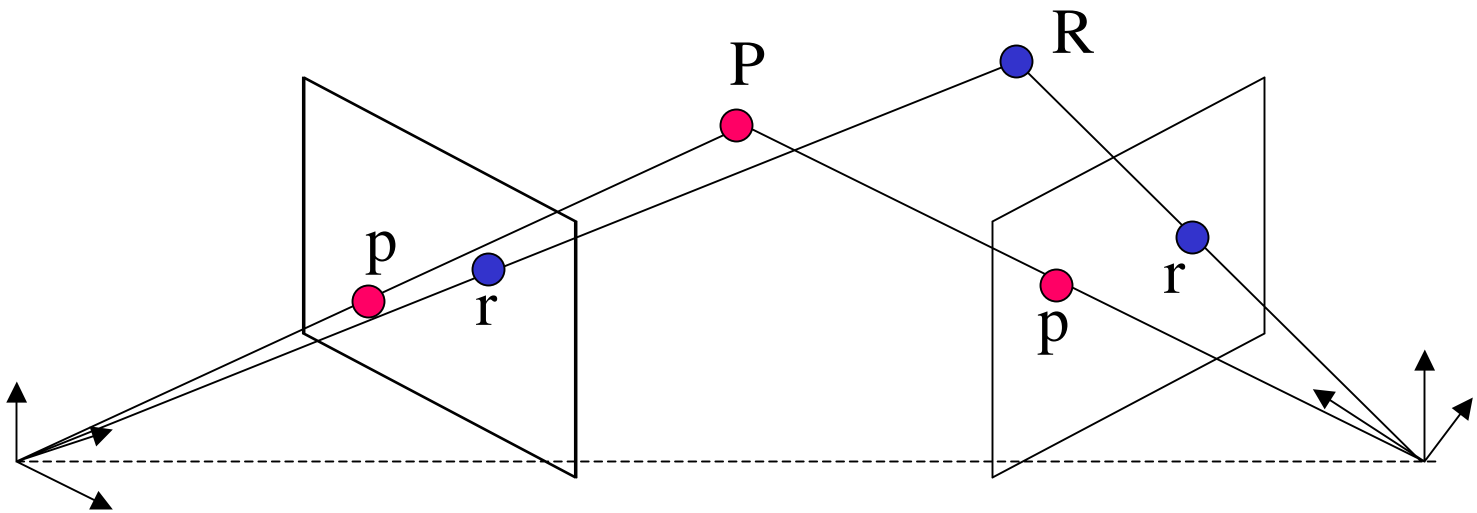
$$X_r = X_r + B \quad (3)$$

$$Z = Z_l = Z_r \quad (4)$$

Z can be determined by equation (8) since:

- **the correspondence is established (red dots) and disparity computed.**
- **the transformation between the left and right camera reference system is known (equations 3 and 4)**
- **Intrinsic camera parameter (f) is known**

A general stereo imaging system



Depth recovering depends on:

• determining which points in one image correspond to points in the other image (**Correspondence problem**)

• determining the geometric relationships for the cameras (**Camera calibration**)

The Correspondence problem

The correspondence problem is a search problem:
Given an element in the left image, we search for the Corresponding element in the right image.
This involves two decisions:

- Which image element to match, and
- Which similarity measure to adopt.

Classification of the correspondence algorithms

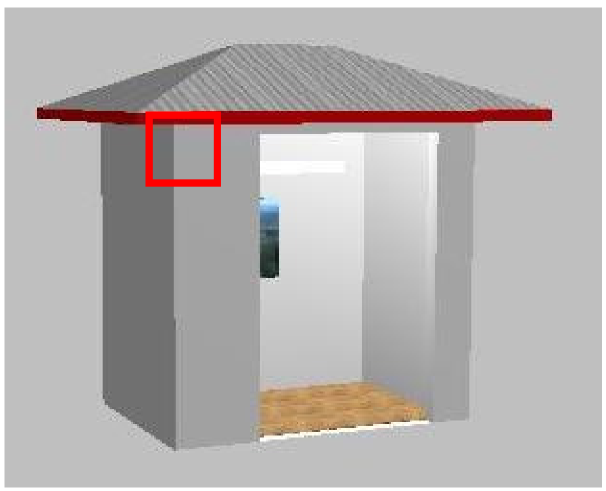
•**Correlation based methods**:
elements to match are Image windows of fixed size and the similarity criterion is a measure of correlation between windows in the two images

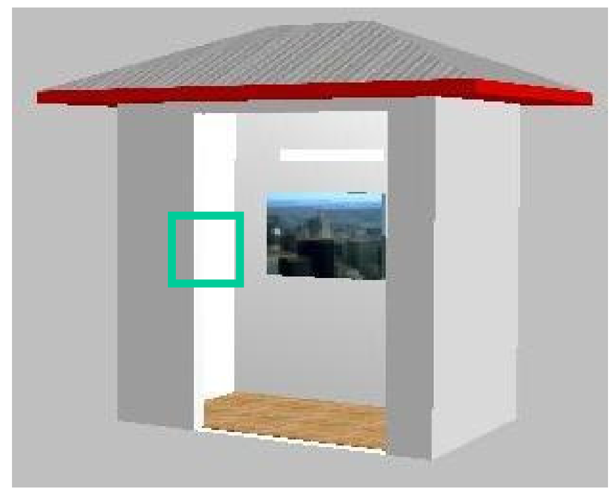
•**Feature based methods**:
Elements to match are set of features extracted from the Image. They use a measure of distance between feature Descriptors. Corresponding elements are given by the most Similar feature pair, the one associated with the minimum distance

Correlation based methods (1)

I_l

Ir





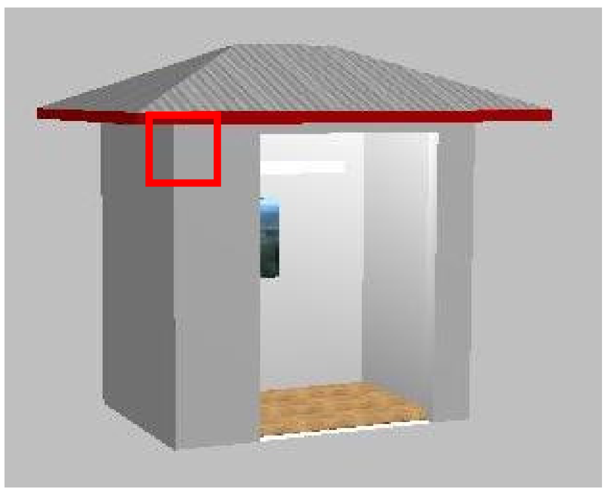
$$C(d_x, d_y) = \sum_{m=-W}^{W} \sum_{n=-W}^{W} (I_l(i+m, j+n) - I_r(i+m-d_x, j+n-d_y))^2$$

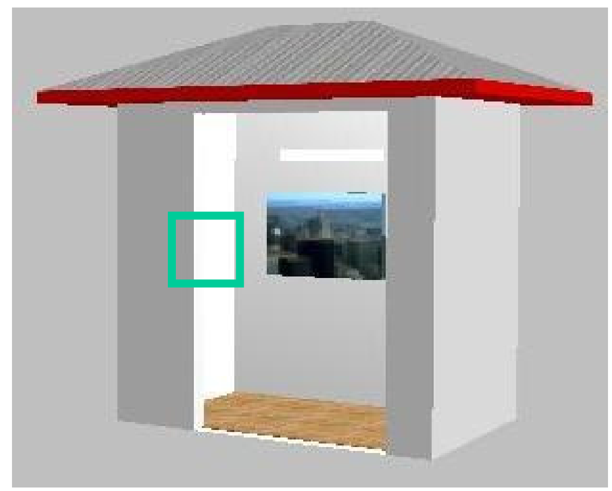
Sum of squared differences

Correlation based methods (2)

I_l

Ir

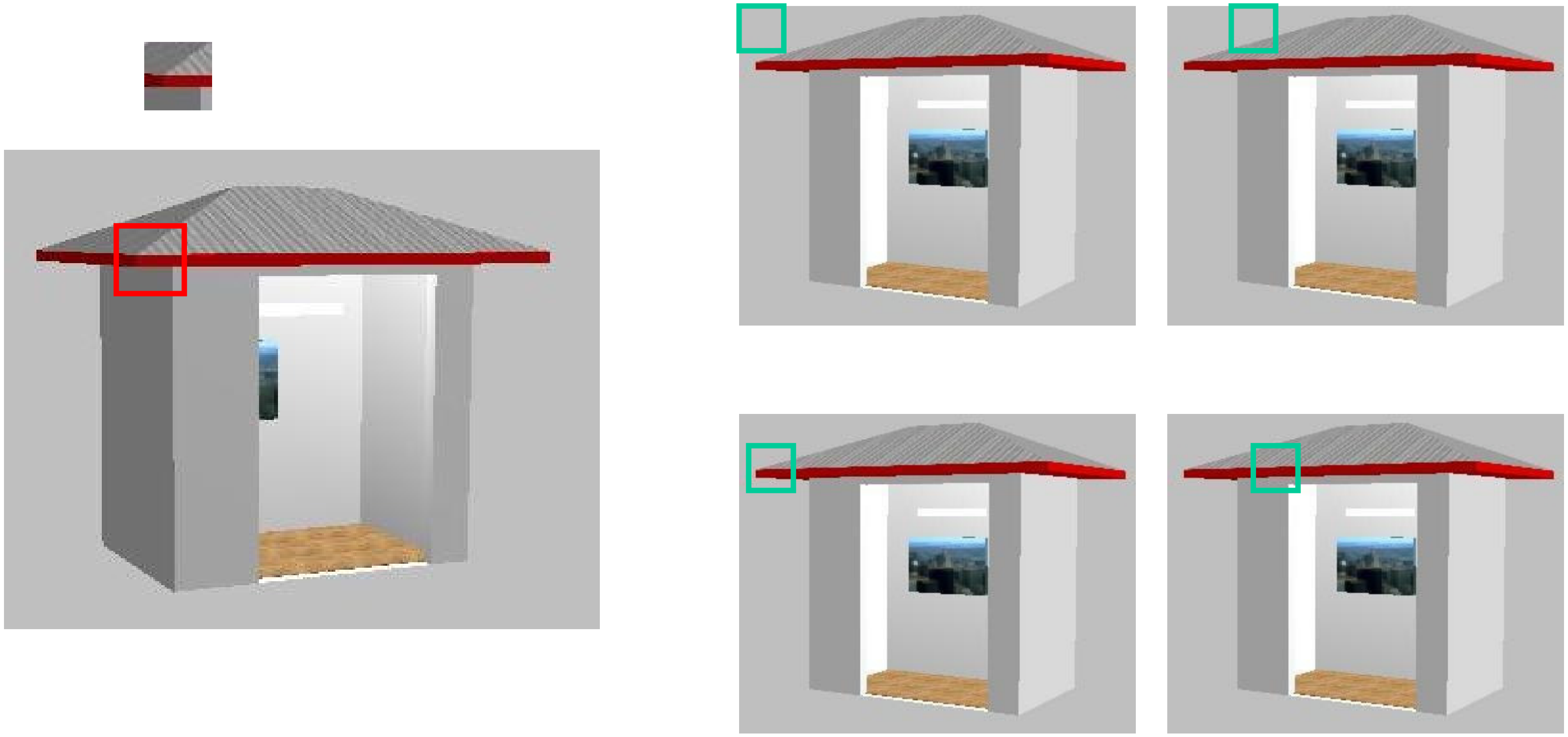




$$C\left(d_x, d_y\right) = \sum_{m=-W}^{W} \sum_{n=-W}^{W} I_l(i+m, j+n)\, I_r(i+m-d_x, j+n-d_y)$$

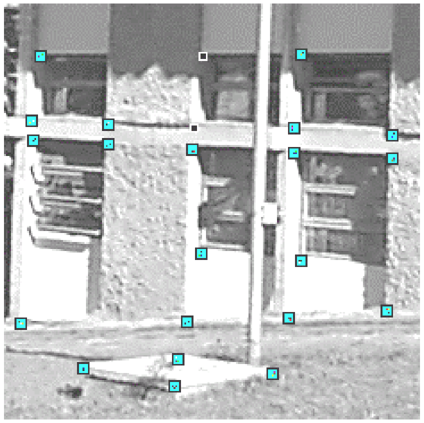
Cross correlation

Correlation based methods (3)



Higher cross correlation
coefficient

Feature based methods (1)




Based on measures of similarity and compatibility of the features (corners, edges, lines, …)

Feature based methods (2)

For the feature-based approach, the matching features must have similar attribute values

feature	Similarity measured in terms of:
corner	• surrounding gray values (SSD, Cross-correlation) • location
lines	•orientation •contrast •coordinates of edge or line’s midpoint •length of line

Feature based methods (3)

Comparing lines

- l_l and l_r: line lengths
- θ_l and θ_r: line orientations
- (x_l,y_l) and (x_r,y_r): midpoints
- c_l and c_r: average contrast along lines
- ω_l ω_θ ω_m ω_c : weights controlling influence

$$S = \frac{1}{w_l(l_l - l_r)^2 + w_\theta(l_{\theta l} - l_{\theta r})^2 + w_m((x_l - x_r)^2 - (y_l - y_r)^2) + w_c(c_l - c_r)^2}$$

larger S → more similar the lines

Correlation x Feature based methods

Correlation methods

- Are easier to implement
- Provide dense disparity maps
- Need textured images to work well
- Due to foreshortening effects and change in illumination it is inadequate for matching of pairs taken from different viewpoints

Feature based methods

- Suitable when a priori information is available
- Provide sparse disparity maps
- Relatively insensitive to illumination changes

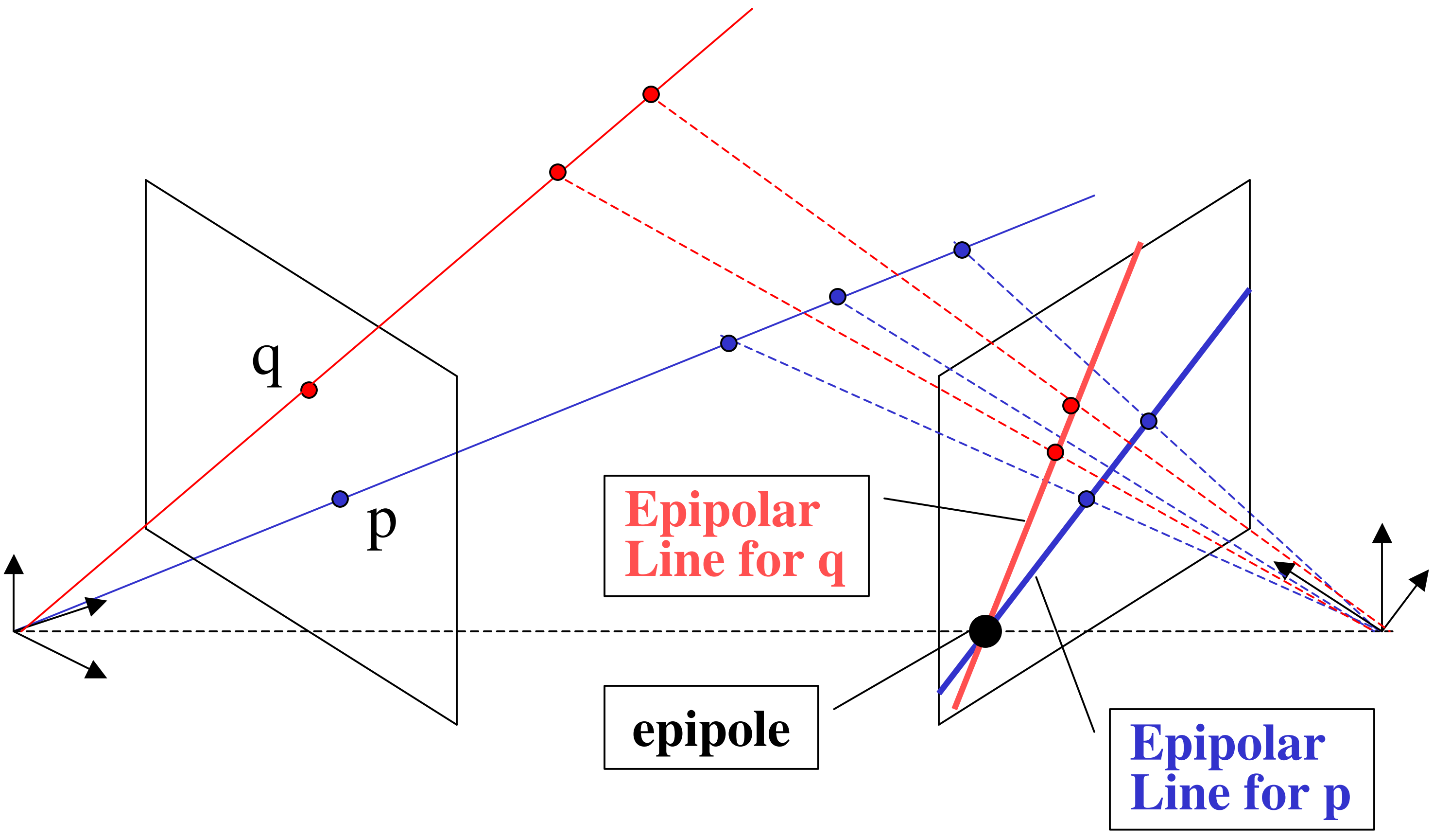
Occlusions and ambiguities – introduce difficulties for both methods

Constraints – reduces the effects of both phenomena

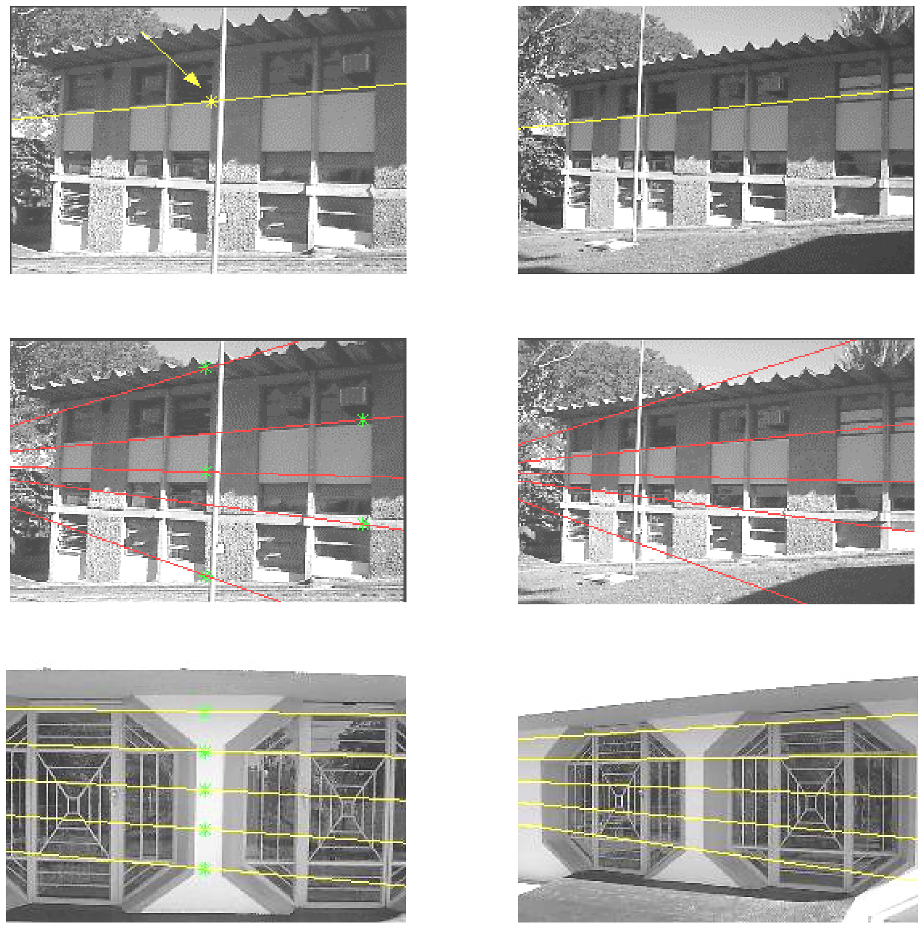
- EPIPOLAR GEOMETRY

The epipolar geometry

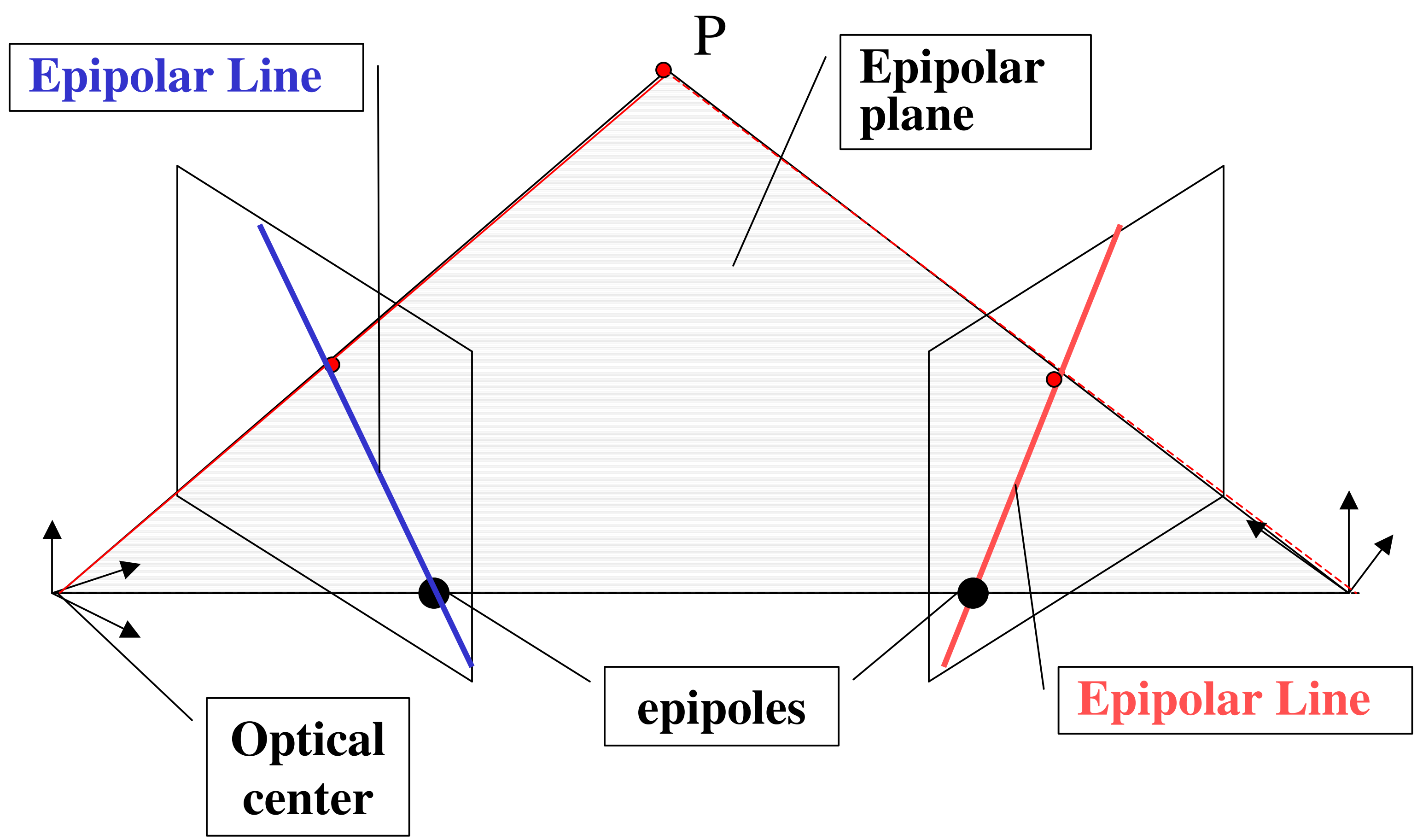
The epipolar geometry



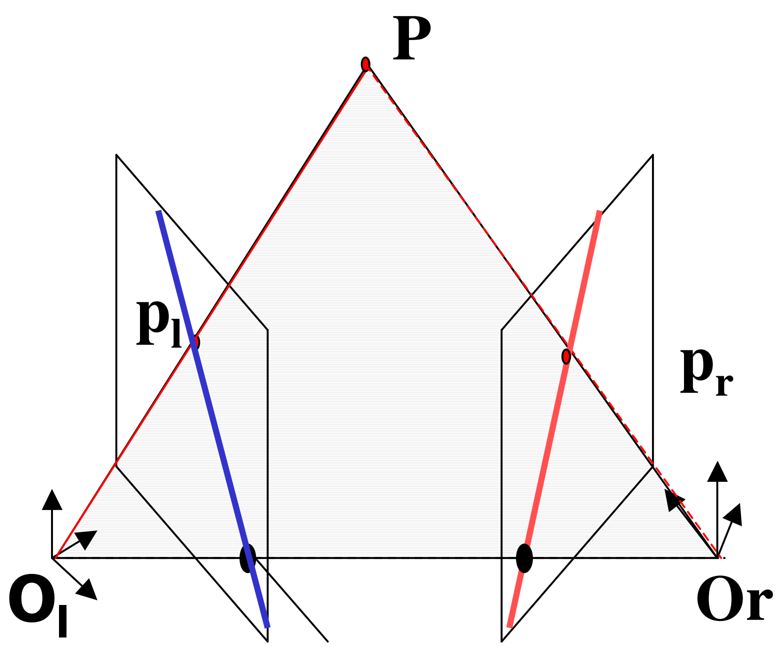
The Epipolar lines



The epipolar plane



The epipolar geometry (notation)



$$\mathbf{T} = (\mathbf{O_r} - \mathbf{O_l})$$

$$\mathbf{P_l} = (X_l, Y_l, Z_l)$$

$$\mathbf{P_r} = (X_r, Y_r, Z_r)$$

Refer to the same 3D point (P) described in the left and right camera coordinate frame

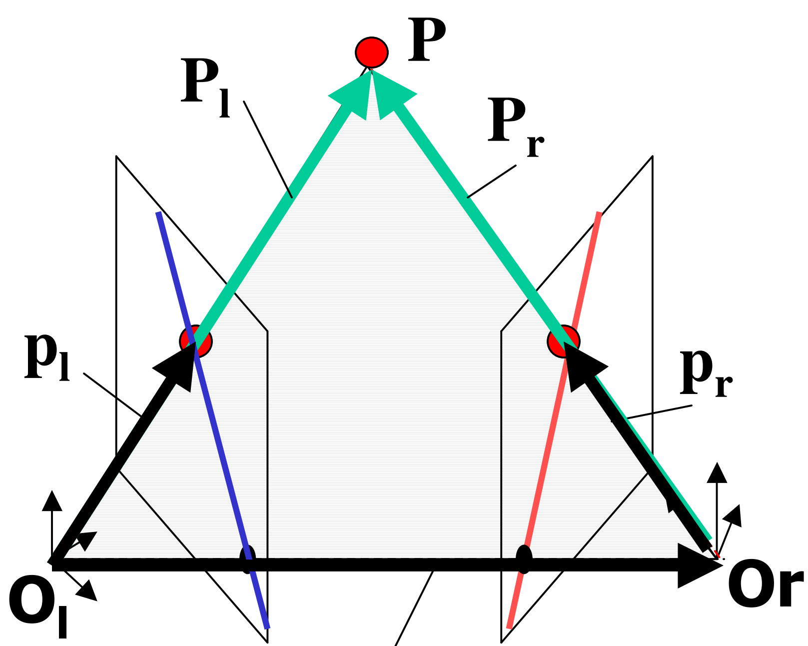
$$\mathbf{p_l} = (x_l, y_l, z_l)$$

$$\mathbf{p_r} = (x_r, y_r, z_r)$$

Refer to the projection of (P) in the left and right camera coordinate frame

$$f_l = z_l \quad \text{and} \quad f_r = z_r$$

Relation between camera reference frames



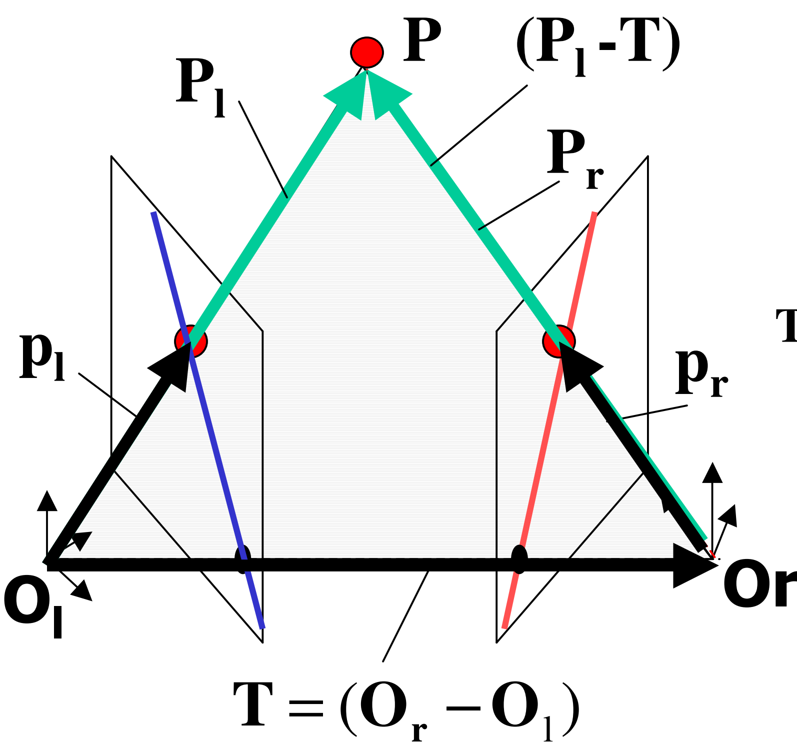
$$\mathbf{T} = (\mathbf{O}_r - \mathbf{O}_l)$$

$$\mathbf{P}_r = R(\mathbf{P}_l - \mathbf{T})$$

$$\mathbf{p}_l = \frac{f_l}{Z_l}$$

$$\mathbf{p}_r = \frac{f_r}{Z_r}$$

The Essential matrix (E)



Given the coplanarity condition of the vectors P_l, T and P_l-T

$$(\mathbf{P_l} - \mathbf{T})^T \mathbf{T} \times \mathbf{P_l} = 0 \quad (1)$$

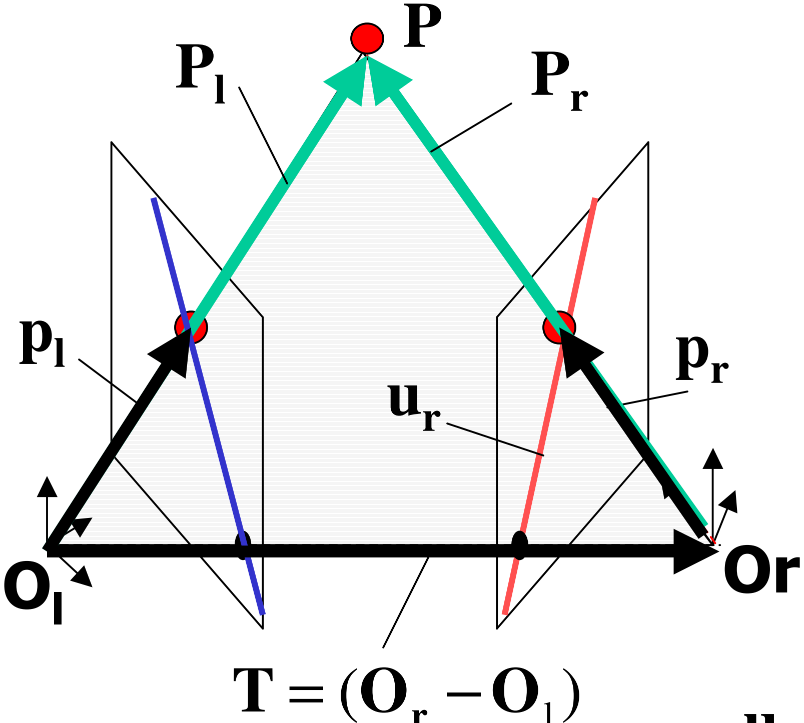
$$\mathbf{T} \times \mathbf{P_l} = \begin{bmatrix} t_y p_z - t_z p_y \\ t_z p_x - t_x p_z \\ t_x p_y - t_y p_x \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}}_{S} \mathbf{P_l} = S\mathbf{P_l} \quad (2)$$

$$(\mathbf{P_l} - \mathbf{T}) = R^T \mathbf{P_r} \quad (3)$$

$$(2)\,(3) \rightarrow (1) \quad \mathbf{P_r}^T E\, \mathbf{P_l} = 0 \quad (4)$$

$$E = RS$$

The Epipolar constraint (1)



$$\mathbf{P_r}^T E \mathbf{P_l} = 0 \quad \rightarrow \quad \mathbf{p_r}^T E\, p_l = 0$$

$$\begin{bmatrix} p_x^r & p_y^r & p_z^r \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{bmatrix} p_x^l \\ p_y^l \\ p_z^l \end{bmatrix} = 0$$

$$a = e_{11}p_x^l + e_{12}p_y^l + e_{13}p_z^l$$
$$b = e_{21}p_x^l + e_{22}p_y^l + e_{23}p_z^l$$
$$c = e_{31}p_x^l + e_{32}p_y^l + e_{33}p_z^l$$

$$a\,p_x^r + b\,p_y^r + c\,f_r = 0 \quad (2)$$

$$\mathbf{u_r} = E\,\mathbf{p_l} = \begin{pmatrix} a \\ b \\ c\,f_r \end{pmatrix}$$ ⇐ Coef. of Epipolar line **Ur**

The Epipolar line and the image reference system

$$\mathbf{u_r} = E\,\mathbf{p_l} = \begin{pmatrix} a \\ b \\ c\,f_r \end{pmatrix}$$

⟸ Epipolar line on Camera Reference system

$\mathbf{u_r}$ and $\mathbf{p_l}$ are defined on the camera reference system (E is computed based on extrinsic parameters R and T) but what we measure in the image are **2D** pixel locations

Transformation from camera coordinate system to image coordinates system is obtained considering the intrinsic camera parameters

The fundamental matrix (F)

M_l and M_r intrinsic camera parameters
$\tilde{\mathbf{p}}_\mathbf{l}$ and $\tilde{\mathbf{p}}_\mathbf{r}$ measured in pixels

$$\mathbf{p_l} = M_l^{-1}\,\tilde{\mathbf{p}}_\mathbf{l} \quad (1) \qquad\qquad \mathbf{p_r}^T E\, p_l = 0 \quad (3)$$

$$\mathbf{p_r} = M_r^{-1}\,\tilde{\mathbf{p}}_\mathbf{r} \quad (2)$$

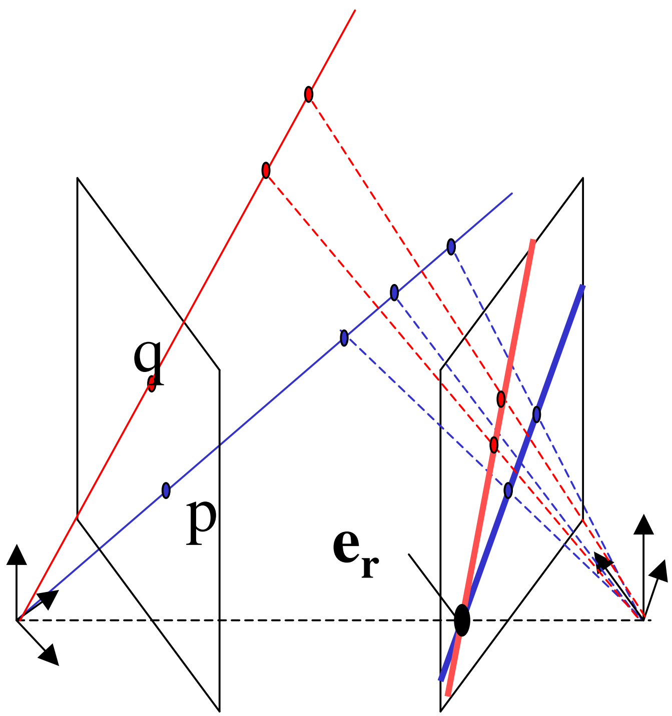
$$(1)\,(2)\,(3) \rightarrow (M_r^{-1}\,\tilde{\mathbf{p}}_\mathbf{r})^T E\,(M_l^{-1}\,\tilde{\mathbf{p}}_\mathbf{l}) = 0$$

$$\tilde{\mathbf{p}}_\mathbf{r}^{\,T}(M_r^{-1})^T E\, M_l^{-1}\,\tilde{\mathbf{p}}_\mathbf{l} = 0 \quad \longrightarrow \quad \tilde{\mathbf{p}}_\mathbf{r}^{\,T} F\, \tilde{\mathbf{p}}_\mathbf{l} = 0$$

$$F = (M_r^{-1})^T E\, M_l^{-1}$$ ⬅ Fundamental matrix

$$\tilde{\mathbf{u}}_\mathbf{r} = F\,\tilde{\mathbf{p}}_\mathbf{l}$$ ⬅ Epipolar line on Image Reference system

Locating the epipoles (1)



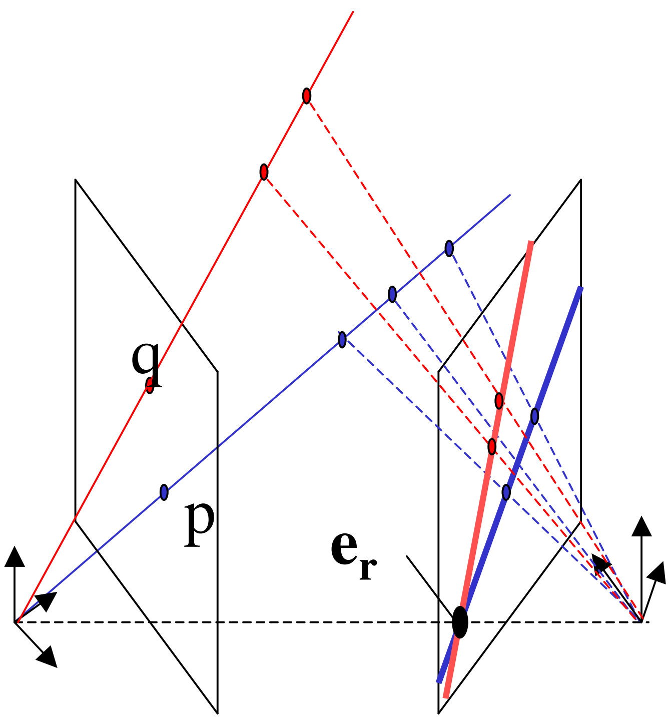
$$\begin{bmatrix} \tilde{p}^r_x & \tilde{p}^r_y & \tilde{p}^r_z \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} \tilde{p}^l_x \\ \tilde{p}^l_y \\ \tilde{p}^l_z \end{bmatrix} = 0$$

For any point $\mathbf{p_l}$ equation (1) is valid
Given $\mathbf{e_r}$ belongs to all epipolar lines

$$\begin{bmatrix} e^r_x & e^r_y & e^r_z \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} \tilde{p}^l_x \\ \tilde{p}^l_y \\ \tilde{p}^l_z \end{bmatrix} = 0 \quad (1)$$

$$\begin{bmatrix} f_{11}e^r_x + f_{21}e^r_y + f_{31}e^r_z \\ f_{12}e^r_x + f_{22}e^r_y + f_{32}e^r_z \\ f_{13}e^r_x + f_{23}e^r_y + f_{33}e^r_z \end{bmatrix} \begin{bmatrix} \tilde{p}^l_x \\ \tilde{p}^l_y \\ \tilde{p}^l_z \end{bmatrix} = 0 \quad (2)$$

Locating the epipoles (2)



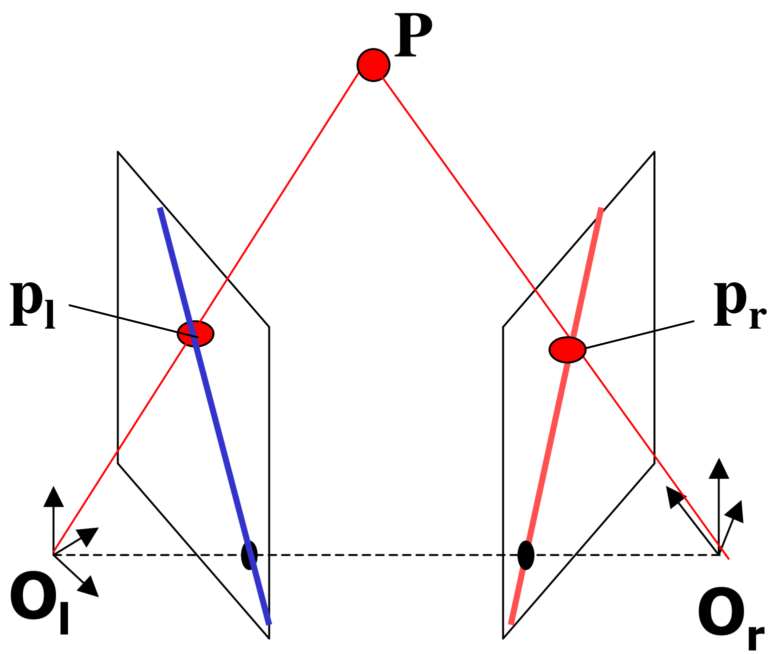
$$\begin{bmatrix} f_{11}e^r_x + f_{21}e^r_y + f_{31}e^r_z \\ f_{12}e^r_x + f_{22}e^r_y + f_{32}e^r_z \\ f_{13}e^r_x + f_{23}e^r_y + f_{33}e^r_z \end{bmatrix} \begin{bmatrix} \tilde{p}^l_x \\ \tilde{p}^l_y \\ \tilde{p}^l_z \end{bmatrix} = 0 \quad (2)$$

Given that $\mathbf{p_l}$ and F are not identically zero, equation (1) holds if and only if

$$\begin{bmatrix} f_{11}e^r_x + f_{21}e^r_y + f_{31}e^r_z \\ f_{12}e^r_x + f_{22}e^r_y + f_{32}e^r_z \\ f_{13}e^r_x + f_{23}e^r_y + f_{33}e^r_z \end{bmatrix} = 0 \quad (3)$$

$$\begin{cases} f_{11}e^r_x + f_{21}e^r_y + f_{31}e^r_z = 0 \\ f_{12}e^r_x + f_{22}e^r_y + f_{32}e^r_z = 0 \\ f_{13}e^r_x + f_{23}e^r_y + f_{33}e^r_z = 0 \end{cases} \Rightarrow F\tilde{\mathbf{e}}_{\mathbf{r}} = 0$$

Computing the Fundamental Matrix (1)



Given a pair of corresponding points in the left and right images equation (1) can be written

$$\begin{bmatrix}\tilde{p}^r_x & \tilde{p}^r_y & \tilde{p}^r_z\end{bmatrix}\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33}\end{bmatrix}\begin{bmatrix}\tilde{p}^l_x \\ \tilde{p}^l_y \\ \tilde{p}^l_z\end{bmatrix} = 0 \quad (1)$$

$$\tilde{p}^r_x(f_{11}\tilde{p}^l_x + f_{12}\tilde{p}^l_y + f_{13}\tilde{p}^l_z) + \tilde{p}^r_y(f_{21}\tilde{p}^l_x + f_{22}\tilde{p}^l_y + f_{23}\tilde{p}^l_z) + \tilde{p}^r_z(f_{31}\tilde{p}^l_x + f_{32}\tilde{p}^l_y + f_{33}\tilde{p}^l_z) = 0 \quad (2)$$

$$\tilde{p}^l_x\tilde{p}^r_x f_{11} + \tilde{p}^l_y\tilde{p}^r_x f_{12} + \tilde{p}^l_z\tilde{p}^r_x f_{13} + \tilde{p}^l_x\tilde{p}^r_y f_{21} + \tilde{p}^l_y\tilde{p}^r_y f_{22} + \tilde{p}^l_z\tilde{p}^r_y f_{23} + \tilde{p}^l_x\tilde{p}^r_z f_{31} + \tilde{p}^l_y\tilde{p}^r_z f_{32} + \tilde{p}^l_z\tilde{p}^r_z f_{33} = 0 \quad (3)$$

Computing the Fundamental Matrix (2)

If n≥ 8 corresponding points are known, a homogeneous system with n equations and 9 unknowns can be written based on equation (3)

$$\tilde{p}^l_x\tilde{p}^r_x f_{11} + \tilde{p}^l_y\tilde{p}^r_x f_{12} + \tilde{p}^l_z\tilde{p}^r_x f_{13} + \tilde{p}^l_x\tilde{p}^r_y f_{21} + \tilde{p}^l_y\tilde{p}^r_y f_{22} + \tilde{p}^l_z\tilde{p}^r_y f_{23} + \tilde{p}^l_x\tilde{p}^r_z f_{31} + \tilde{p}^l_y\tilde{p}^r_z f_{32} + \tilde{p}^l_z\tilde{p}^r_z f_{33} = 0 \quad (3)$$

$$\begin{bmatrix} \tilde{p}^l_{x(1)}\tilde{p}^r_{x(1)} & \tilde{p}^l_{y(1)}\tilde{p}^r_{x(1)} & \tilde{p}^l_{z(1)}\tilde{p}^r_{x(1)} & \tilde{p}^l_{x(1)}\tilde{p}^r_{y(1)} & \tilde{p}^l_{y(1)}\tilde{p}^r_{y(1)} & \tilde{p}^l_{z(1)}\tilde{p}^r_{y(1)} & \tilde{p}^l_{x(1)}\tilde{p}^r_{z(1)} & \tilde{p}^l_{y(1)}\tilde{p}^r_{z(1)} & \tilde{p}^l_{z(1)}\tilde{p}^r_{z(1)} \\ \tilde{p}^l_{x(2)}\tilde{p}^r_{x(1)} & \tilde{p}^l_{y(2)}\tilde{p}^r_{x(2)} & \tilde{p}^l_{z(2)}\tilde{p}^r_{x(2)} & \tilde{p}^l_{x(2)}\tilde{p}^r_{y(2)} & \tilde{p}^l_{y(1)}\tilde{p}^r_{y(2)} & \tilde{p}^l_{z(2)}\tilde{p}^r_{y(2)} & \tilde{p}^l_{x(2)}\tilde{p}^r_{z(2)} & \tilde{p}^l_{y(v)}\tilde{p}^r_{z(2)} & \tilde{p}^l_{z(2)}\tilde{p}^r_{z(v)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{p}^l_{x(n)}\tilde{p}^r_{x(n)} & \tilde{p}^l_{y(n)}\tilde{p}^r_{x(n)} & \tilde{p}^l_{z(n)}\tilde{p}^r_{x(n)} & \tilde{p}^l_{x(n)}\tilde{p}^r_{y(n)} & \tilde{p}^l_{y(n)}\tilde{p}^r_{y(n)} & \tilde{p}^l_{z(n)}\tilde{p}^r_{y(n)} & \tilde{p}^l_{x(n)}\tilde{p}^r_{z(n)} & \tilde{p}^l_{y(n)}\tilde{p}^r_{z(n)} & \tilde{p}^l_{z(n)}\tilde{p}^r_{z(n)} \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0$$

Computing the Fundamental Matrix (3)

System of equations has a unique solution up to a scaling factor

$$f_{33} = 1$$

$$\begin{bmatrix}
\tilde{p}^l_{x(1)}\tilde{p}^r_{x(1)} & \tilde{p}^l_{y(1)}\tilde{p}^r_{x(1)} & \tilde{p}^l_{z(1)}\tilde{p}^r_{x(1)} & \tilde{p}^l_{x(1)}\tilde{p}^r_{y(1)} & \tilde{p}^l_{y(1)}\tilde{p}^r_{y(1)} & \tilde{p}^l_{z(1)}\tilde{p}^r_{y(1)} & \tilde{p}^l_{x(1)}\tilde{p}^r_{z(1)} & \tilde{p}^l_{y(1)}\tilde{p}^r_{z(1)} \\
\tilde{p}^l_{x(2)}\tilde{p}^r_{x(1)} & \tilde{p}^l_{y(2)}\tilde{p}^r_{x(2)} & \tilde{p}^l_{z(2)}\tilde{p}^r_{x(2)} & \tilde{p}^l_{x(2)}\tilde{p}^r_{y(2)} & \tilde{p}^l_{y(1)}\tilde{p}^r_{y(2)} & \tilde{p}^l_{z(2)}\tilde{p}^r_{y(2)} & \tilde{p}^l_{x(2)}\tilde{p}^r_{z(2)} & \tilde{p}^l_{y(v)}\tilde{p}^r_{z(2)} \\
. & . & . & . & . & . & . & . \\
. & & & & & & & . \\
. & & & & & & & . \\
. & & & & & & & . \\
. & & & & & & & . \\
. & . & . & . & . & . & . & . \\
\tilde{p}^l_{x(n)}\tilde{p}^r_{x(n)} & \tilde{p}^l_{y(n)}\tilde{p}^r_{x(n)} & \tilde{p}^l_{z(n)}\tilde{p}^r_{x(n)} & \tilde{p}^l_{x(n)}\tilde{p}^r_{y(n)} & \tilde{p}^l_{y(n)}\tilde{p}^r_{y(n)} & \tilde{p}^l_{z(n)}\tilde{p}^r_{y(n)} & \tilde{p}^l_{x(n)}\tilde{p}^r_{z(n)} & \tilde{p}^l_{y(n)}\tilde{p}^r_{z(n)}
\end{bmatrix}
\begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \end{bmatrix}
=
\begin{bmatrix} \tilde{p}^l_{z(1)}\tilde{p}^r_{z(1)} \\ \tilde{p}^l_{z(2)}\tilde{p}^r_{z(v)} \\ \\ \\ \\ \\ \tilde{p}^l_{z(n)}\tilde{p}^r_{z(n)} \end{bmatrix}$$

Computing the Fundamental Matrix (4)

- Fundamental matrix is unstable when the scene is close to planarity
- Fundamental matrix is unstable when the camera motion is close to pure rotation
- Switching between models can produce discontinuous estimations
- The estimation of homography matrices is always possible

The eight point algorithm

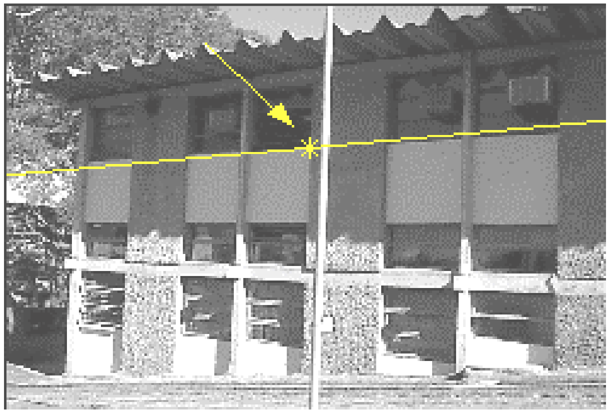
The eight point algorithm (as seen) has the advantage of simplicity of implementation. However, it is extremely susceptible to noise and hence virtually useless for most purposes. In the paper:

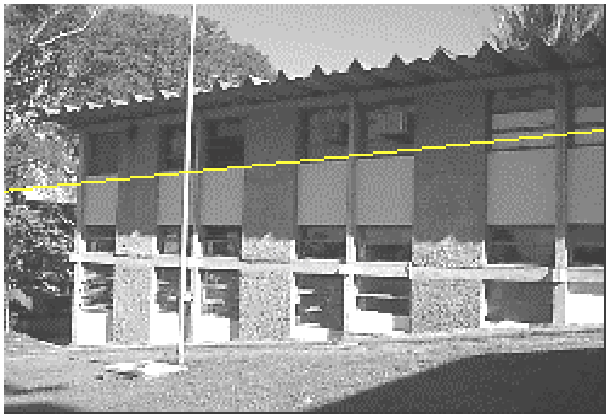
- "In defense of the eight point algorithm", Richard I. Hartley, **IEEE Transactions On Pattern Analysis And Machine Intelligence**, Vol. 19, No. 6, June 1997.

It is shown that by preceding the algorithm with a very simple normalization (translation and scaling) of the coordinates of the matched points, results are obtained comparable with the best iterative algorithms.

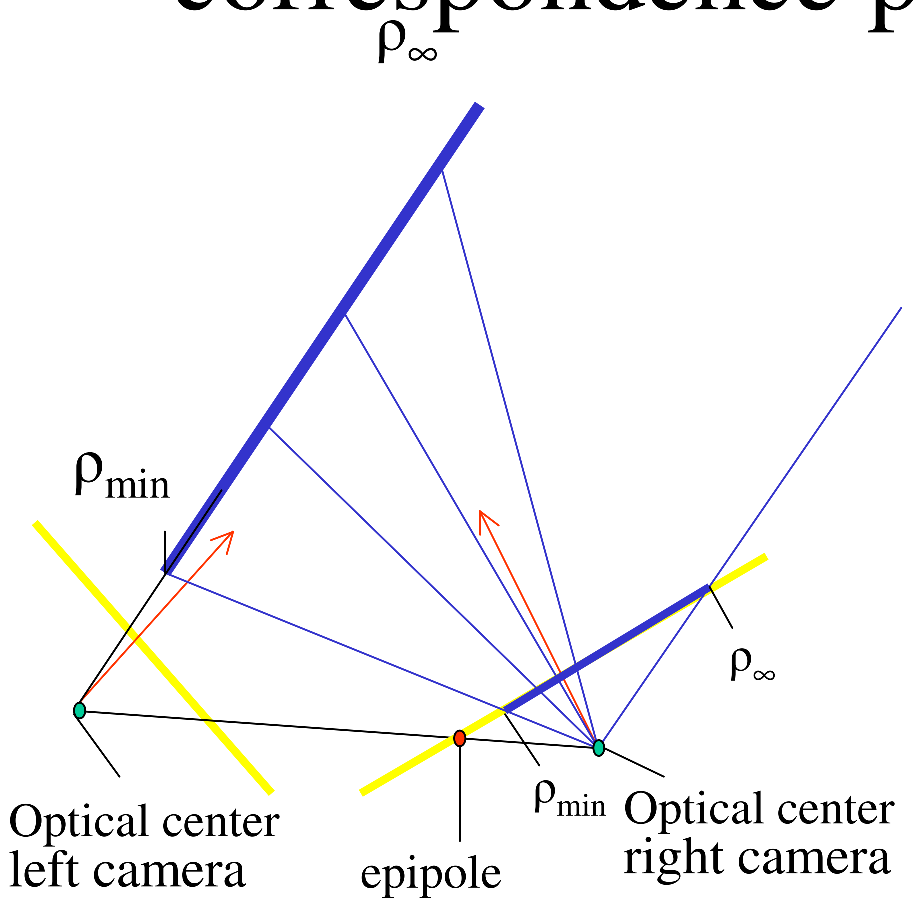
Epipolar constraint and the correspondence problem (1)

Epipolar Constraint Reduces correspondence problem to 1D search along **conjugate epipolar lines**





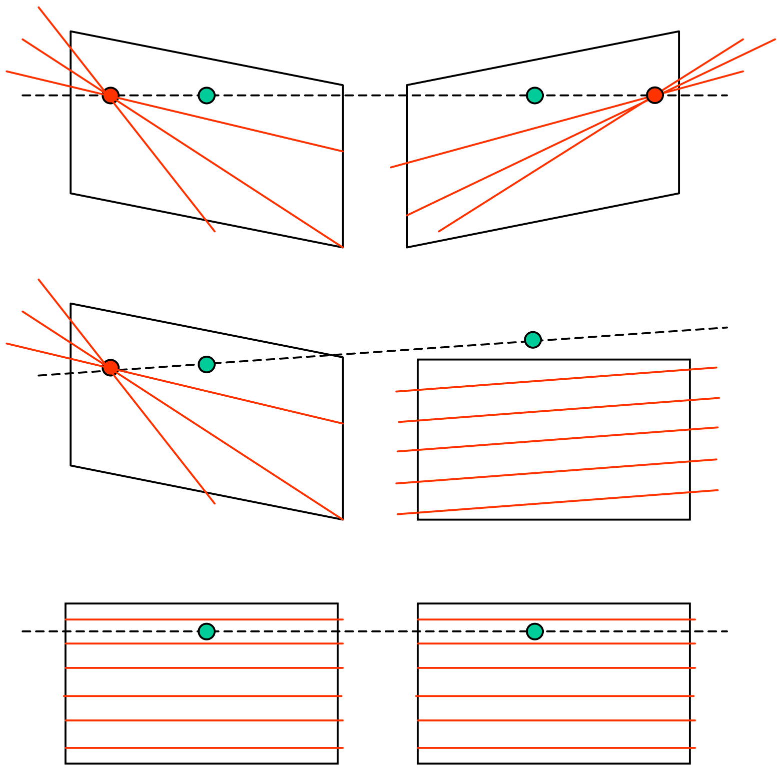
Epipolar constraint and the correspondence problem (2)



Most conventional stereo algorithms contain a fixed limit on the size of the epipolar search bands.

This constraints is usually applied because the data in the scene is known to exist over a limited range of depths.

The bundle of epipolar lines

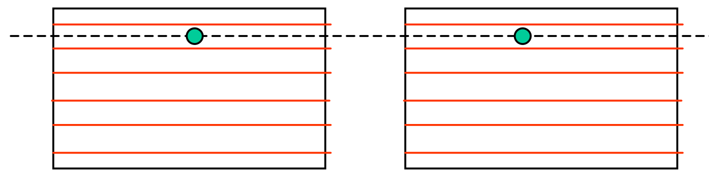


General case

Parallel epipolar lines
in one of the images

Parallel epipolar lines
in both images

Correspondence: simplest case



- Image planes of cameras are parallel
- Focal points are at same height
- Focal lengths same
- Then, epipolar lines are horizontal scan lines

It is always possible to achieve this geometry with image rectification

Image rectification

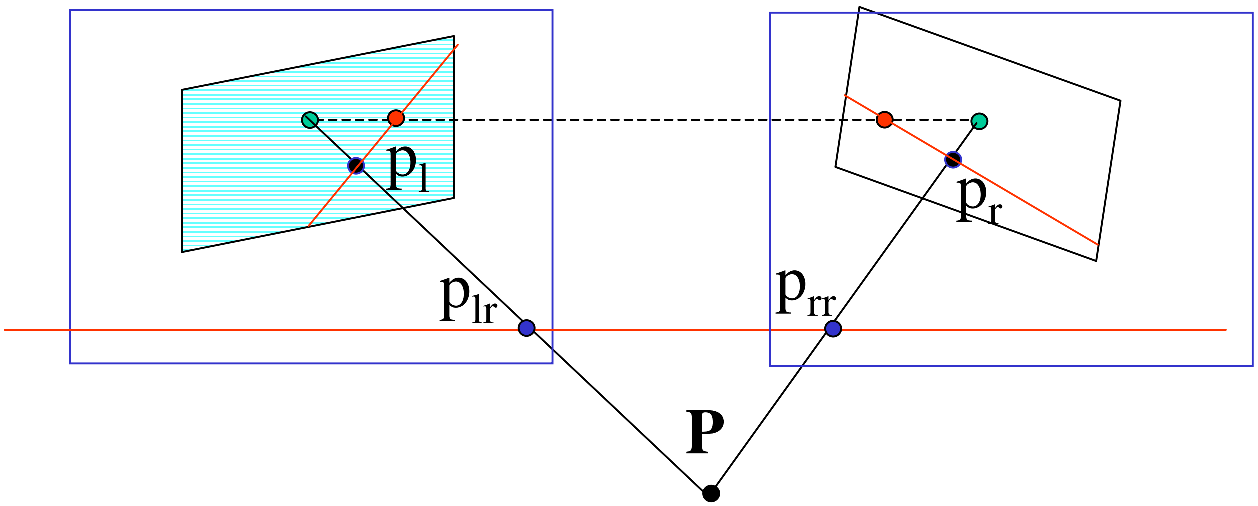
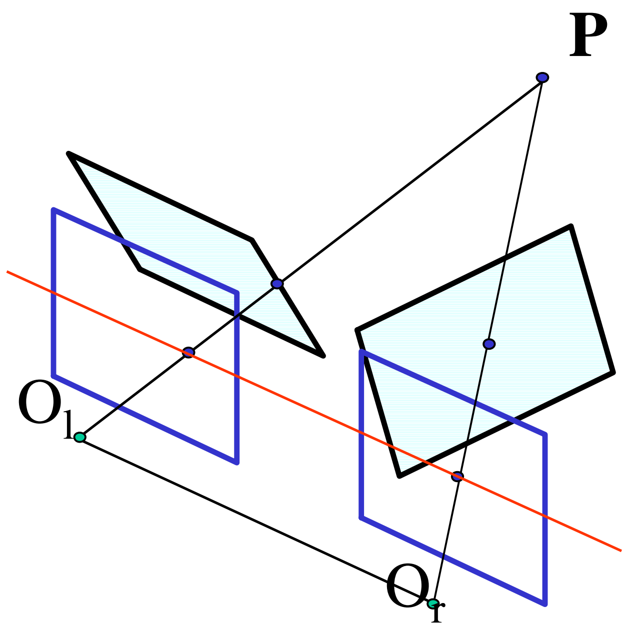


Image rectification (1)



Re-project image planes onto common plane parallel to line between optical centers to create a new stereo image pair such that the epipolar lines are horizontal and identical for the 2 new images

Image rectification (example)

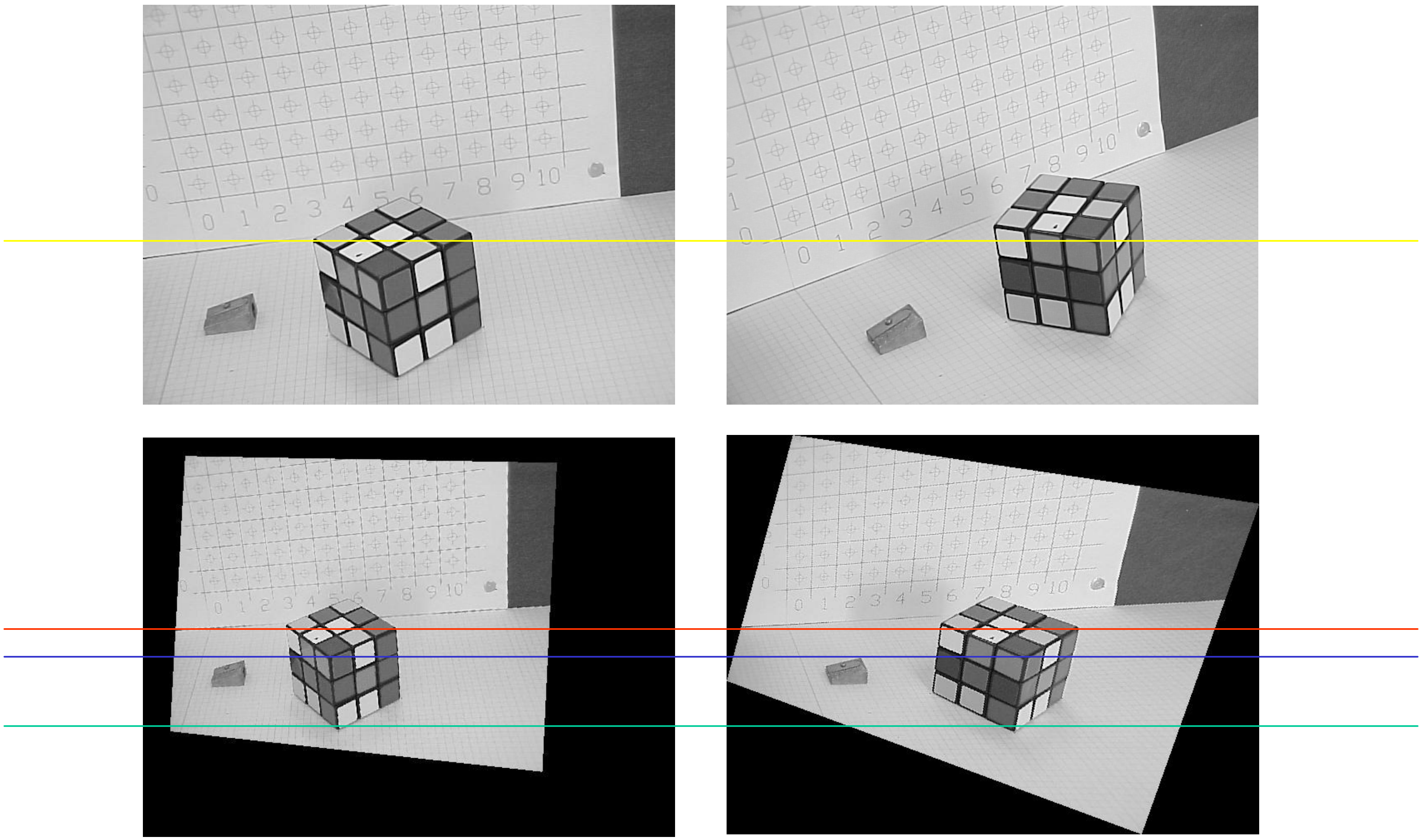
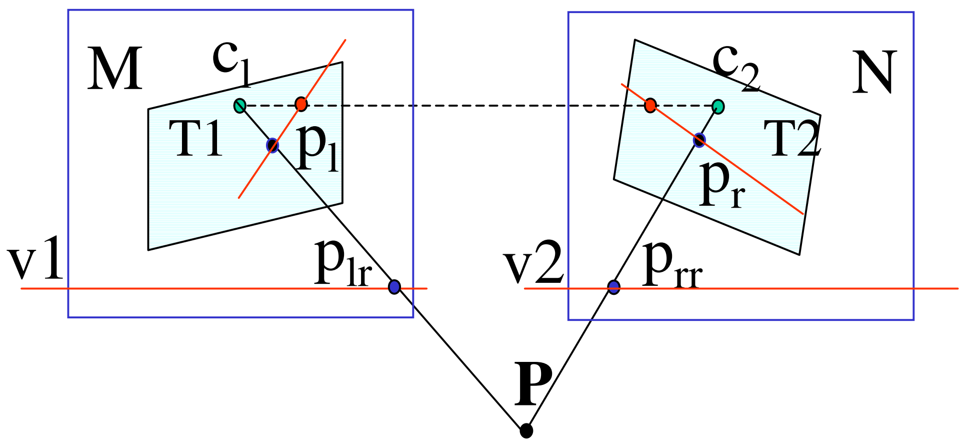


Image rectification (problem - 1)



Problem: given the camera models T1 and T2 obtain the new camera models M and N considering the constraints:

- The optical centers of M and N are C1 and C2 respectively (to give a unique match between points P_{lr} and P_l respectively before and after rectification)
- The focal plane of M is identified with that of N (to produce parallel lines in both images)
- For any point P (not in optical plane), the y coordinates of the image points P_{lr} and P_{rr} obtained by M and N respectively are equal (v1=v2)

Image rectification (solution)

(Artificial vision for mobile robots, Nicholas Ayache. MIT Press, 1991 – section 3.6)

$$M = \begin{vmatrix} [(C_1 x\, C_2)\, x\, C_1]^T & 0 \\ (C_1 x\, C_2)^T & 0 \\ [(C_1 - C_2)\, x\, (C_1 x\, C_2)]^T & \|C_1 x\, C_2\|^2 \end{vmatrix} \qquad N = \begin{vmatrix} [(C_1 x\, C_2)\, x\, C_2]^T & 0 \\ (C_1 x\, C_2)^T & 0 \\ [(C_1 - C_2)\, x\, (C_1 x\, C_2)]^T & \|C_1 x\, C_2\|^2 \end{vmatrix}$$

Where C_1 and C_2 are the optical centers of cameras 1 and 2 respectively. The optical center Ci is determined solving the equation system:

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = T_i \begin{pmatrix} x_{ci} \\ y_{ci} \\ z_{ci} \\ 1 \end{pmatrix}$$

where

$$(C_i) = \begin{pmatrix} x_{ci} \\ y_{ci} \\ z_{ci} \end{pmatrix}$$

Coordinates of the camera optical center

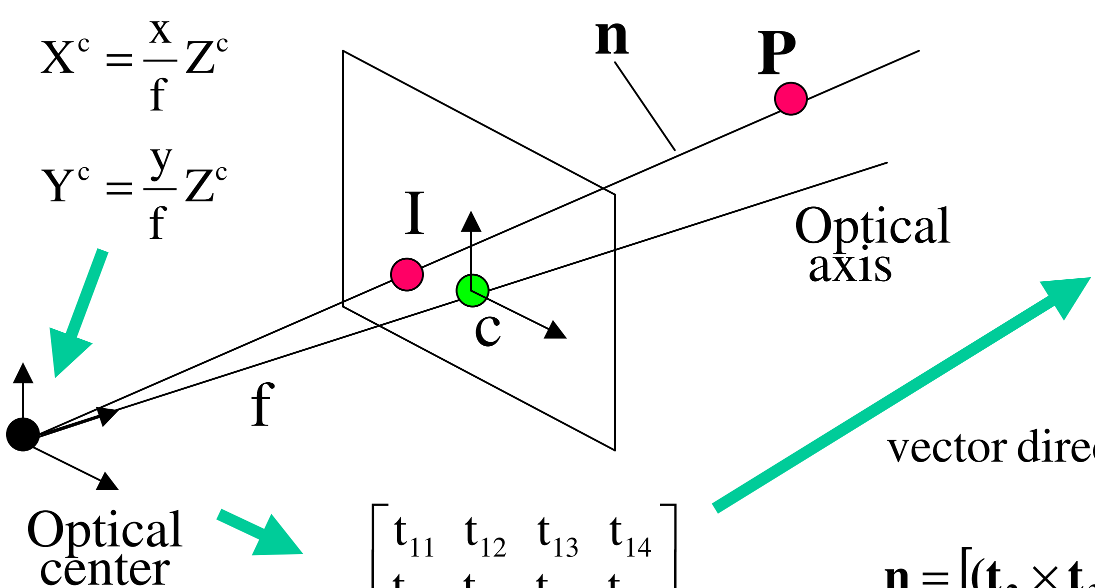
$$T_i = \begin{vmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{34} & t_{34} & t_{34} \end{vmatrix}$$

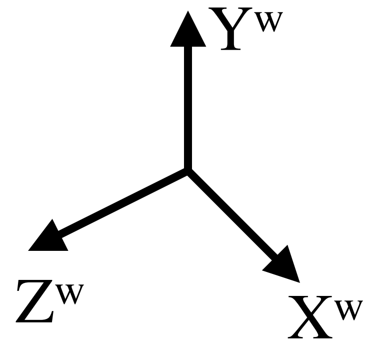
Camera model

The inverse perspective reference system (review-1)

$$X^c = \frac{x}{f} Z^c$$

$$Y^c = \frac{y}{f} Z^c$$





$$\begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix}$$

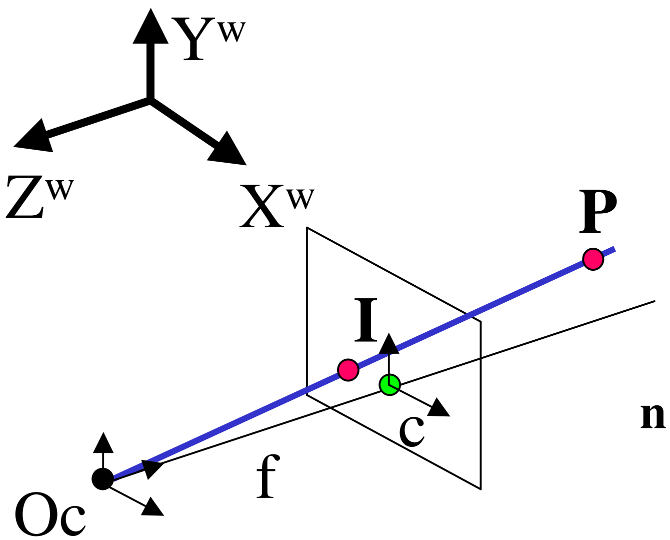
camera projection matrix

vector direction in WRS

$$\mathbf{n} = \left[(\mathbf{t_2} \times \mathbf{t_3}) \;\; (\mathbf{t_3} \times \mathbf{t_1}) \;\; (\mathbf{t_1} \times \mathbf{t_2}) \right] \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = S \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

where $t_i = \left[t_{i1} \;\; t_{i2} \;\; t_{i3} \right]^T$

The inverse perspective reference system (review-2)



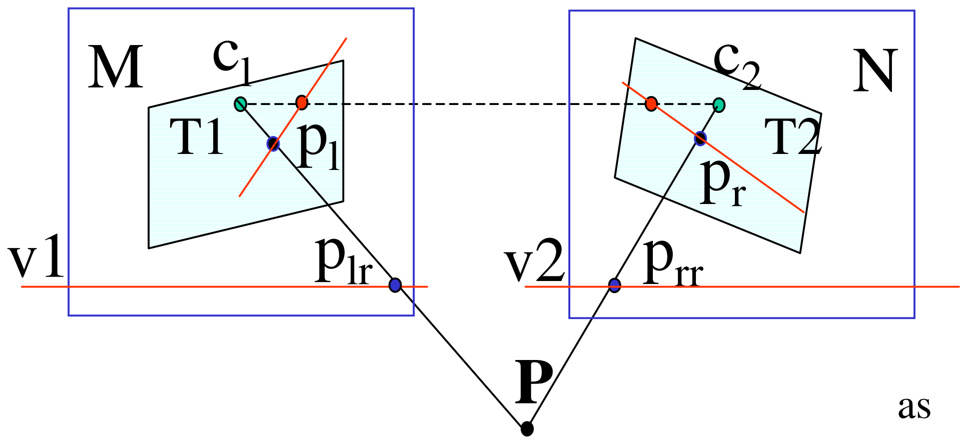
$$\mathbf{n} = \left[(\mathbf{t_2} \times \mathbf{t_3}) \; (\mathbf{t_3} \times \mathbf{t_1}) \; (\mathbf{t_1} \times \mathbf{t_2}) \right] \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = S \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$S = \begin{bmatrix} (t_{22}t_{33} - t_{32}t_{23}) & (t_{32}t_{13} - t_{12}t_{33}) & (t_{12}t_{23} - t_{22}t_{13}) \\ (t_{31}t_{23} - t_{21}t_{33}) & (t_{11}t_{33} - t_{31}t_{13}) & (t_{21}t_{13} - t_{11}t_{23}) \\ (t_{21}t_{32} - t_{31}t_{22}) & (t_{31}t_{12} - t_{11}t_{32}) & (t_{11}t_{22} - t_{21}t_{12}) \end{bmatrix}$$

$$\mathbf{n} = \begin{bmatrix} x(t_{22}t_{33} - t_{32}t_{23}) + y(t_{32}t_{13} - t_{12}t_{33}) + (t_{12}t_{23} - t_{22}t_{13}) \\ x(t_{31}t_{23} - t_{21}t_{33}) + y(t_{11}t_{33} - t_{31}t_{13}) + (t_{21}t_{13} - t_{11}t_{23}) \\ x(t_{21}t_{32} - t_{31}t_{22}) + y(t_{31}t_{12} - t_{11}t_{32}) + (t_{11}t_{22} - t_{21}t_{12}) \end{bmatrix}$$

The parametric equation of the line OcI in WRS is: $\mathbf{P} = \mathbf{Oc} + \lambda \mathbf{n}$

Image rectification (computation)



P_{lr} can be computed determining the inverse perspective for pl (from T1) $\mathbf{P} = \mathbf{Oc} + \lambda \mathbf{n}$

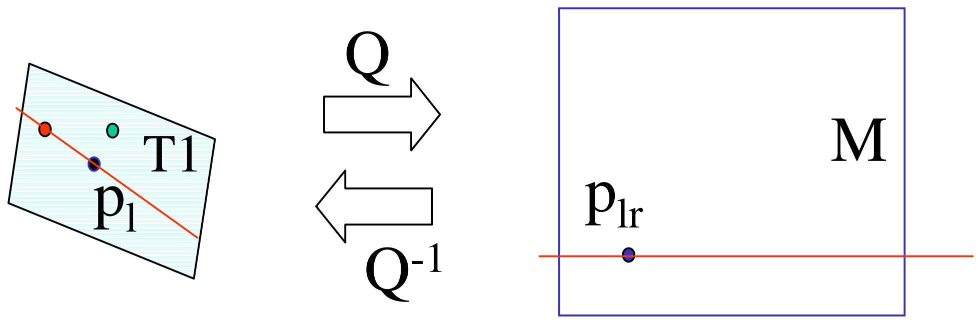
and re-projecting one point in the line using the new matrix M

$$\mathbf{p}_{lr} = M \begin{pmatrix} \mathbf{Oc} + \mathbf{n} \\ 1 \end{pmatrix} \qquad (1)$$

as $M \begin{pmatrix} \mathbf{Oc} \\ 1 \end{pmatrix} = 0 \rightarrow$ (1) can be written

$$\mathbf{p}_{lr} = \begin{vmatrix} [(C_1 \times C_2) \times C_1]^T \\ (C_1 \times C_2)^T \\ [(C_1 - C_2) \times (C_1 \div C_2)]^T \end{vmatrix} \left| (t_2 \times t_3) \ (t_3 \times t_1) \ (t_1 \times t_2) \right| (\mathbf{p}_l)$$

Image rectification (computation)



$$\mathbf{p}_{lr} = Q \begin{vmatrix} x_1 \\ y_1 \\ 1 \end{vmatrix} \qquad Q = \begin{vmatrix} [(C_1 x\ C_2)\ x\ C_1]^T \\ (C_1 x\ C_2)^T \\ [(C_1 - C_2)\ x\ (C_1 x\ C_2)]^T \end{vmatrix} \begin{vmatrix} (t_2\ x\ t_3) & (t_3\ x\ t_1) & (t_1\ x\ t_2) \end{vmatrix}$$

Rectified coordinates are in general not integers. To obtain integer coordinates rectification is implemented backwards applying inverse transformation (Q^{-1}) and pixels values in the new image computed as bilinear interpolation of pixels in the old image

Image rectification

(other approaches)

Problem 1 : given the intrinsic parameters of each camera and the extrinsic parameters of the stereo system (R and T) obtain the transform for image rectification.

- Introductory techniques for 3D Computer Vision, Emanuele Trucco e Alessandro Verri. Prentice Hall. 1998, section 7.3

Problem 2 : given the Fundamental matrix (F) of the stereo system obtain the transform for image rectification.

- Zezhi Chen et all., A new image rectification algorithm. Pattern Recognition Letters 24 (2003) 251–260
- Forster, Carlos Henrique Quartucci, Alinhamento Imagem-Modelo Baseada na Visão Estéreode Regiões Planares Arbitrárias. Tese de doutorado. FEEC-UNICAMP, 2004

3D Reconstruction

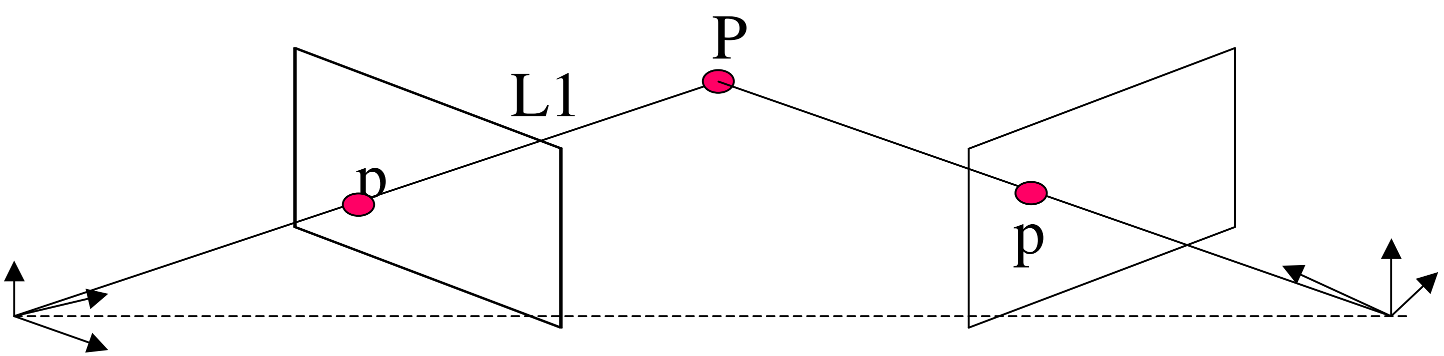
3D Reconstruction

The 3D reconstruction that can be obtained depends on the *a priori* knowledge available on the parameters of the stereo system.
Three cases are identified:

- Intrinsic and extrinsic parameters are known
- Only intrinsic parameters are known
- Neither the intrinsic nor the extrinsic parameters are known

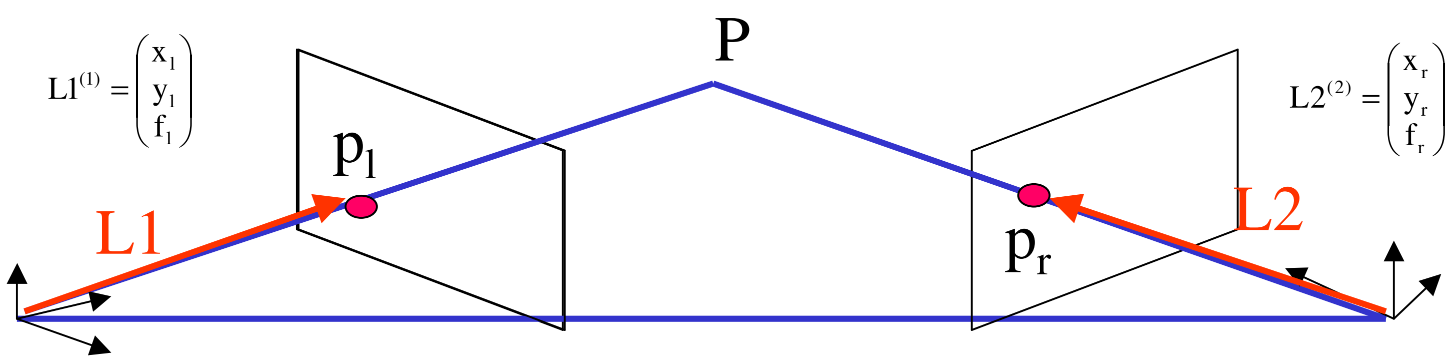
Reconstruction for intrinsically and extrinsically calibrated cameras

Under assumption that the intrinsic and extrinsic parameters are known, the point P is computed from its projections on left and right camera by the intersection of lines L1 and L2



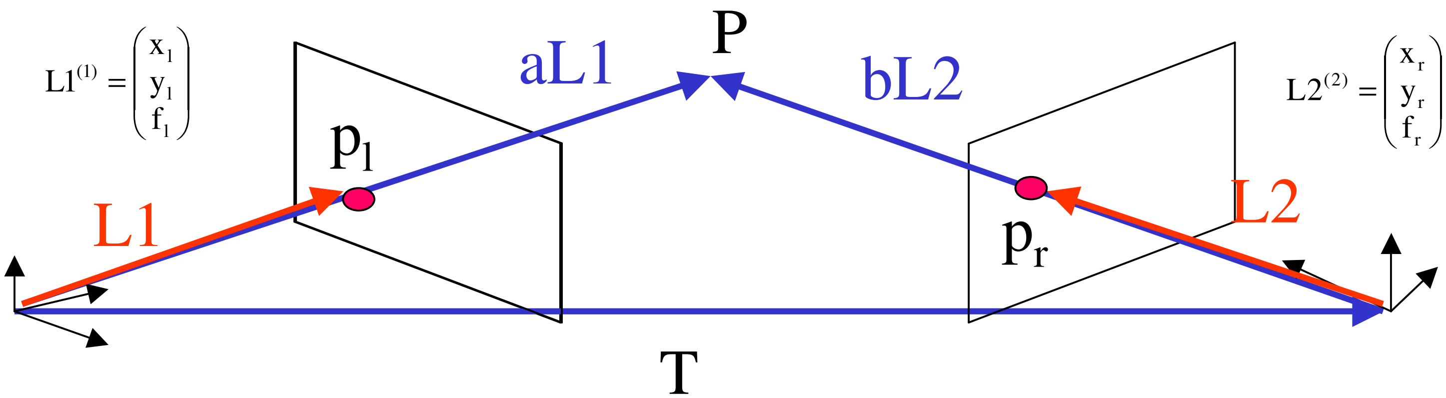
Reconstruction by triangulation (1)

• Vectors L1 and L2 can be computed from the image points Pl and Pr respectively (intrinsic parameters are known).
• Vectors L1 and L2 are expressed in their respective reference frames
• To compute the intersection vectors L1 and L2 must be expressed in the same reference frame



Reconstruction by triangulation (2)

- To express L2 into the same reference as L1, L2 must be rotated by R^{-1} and translated by T (where R and T are the extrinsic parameters of the stereo system)



$$\mathbf{P_r} = \mathbf{R}(\mathbf{P_l} - \mathbf{T}) \longrightarrow \mathbf{L2}^{(1)} = (\mathbf{R}^{-1}\,\mathbf{L2}^{(2)}) + \mathbf{T}$$

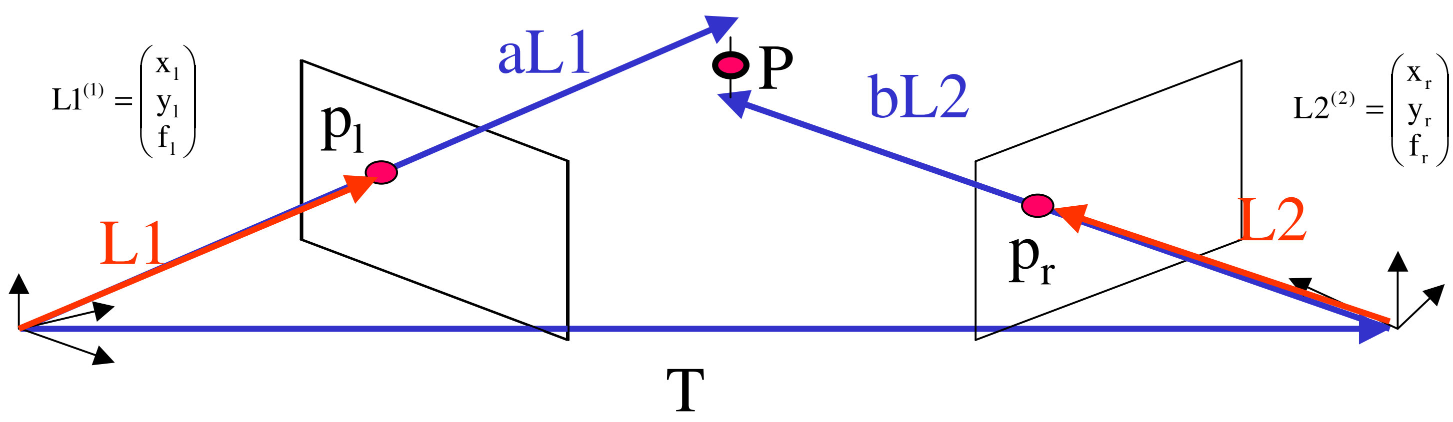
$$a\,\mathbf{L1} - \left[(b\,\mathbf{R}^{-1}\,\mathbf{L2}^{(2)}) + \mathbf{T}\right] = \mathbf{T}$$

$$a\,\mathbf{L1} - b\,\mathbf{R}^{-1}\,\mathbf{L2}^{(2)} = \mathbf{T}$$

⇐ Linear system of equations

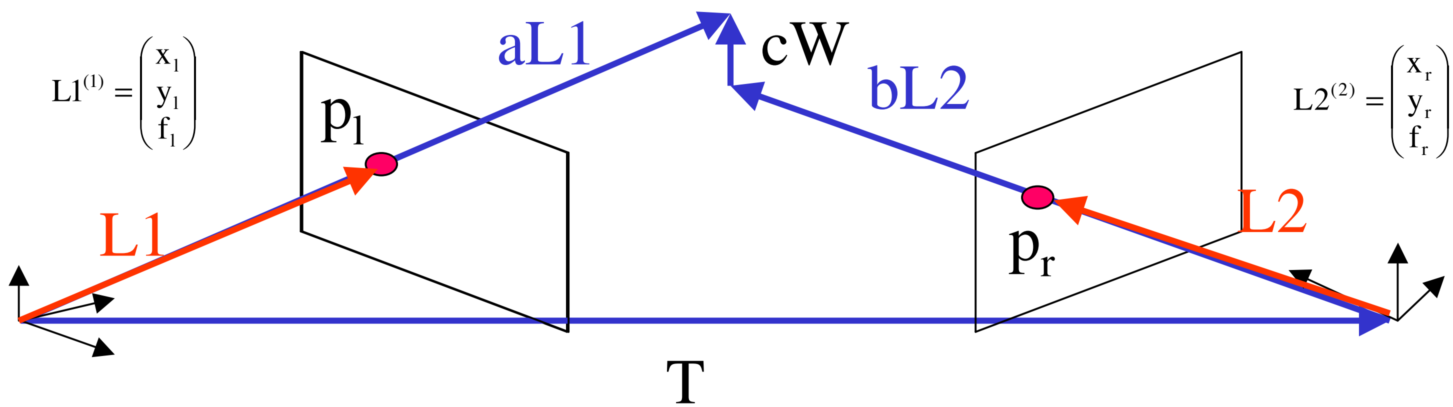
Reconstruction by triangulation (3)

- Since parameters and image locations are known only approximately , the two rays will not actually intersect in the space
- Their intersection can only be estimated as the point of minimum distance of both rays



Reconstruction by triangulation (4)

- If vector **W** is assumed to be perpendicular to L1 and L2, the problem reduces to determining the midpoint of the segment parallel to vector **W** that joins L1 and L2



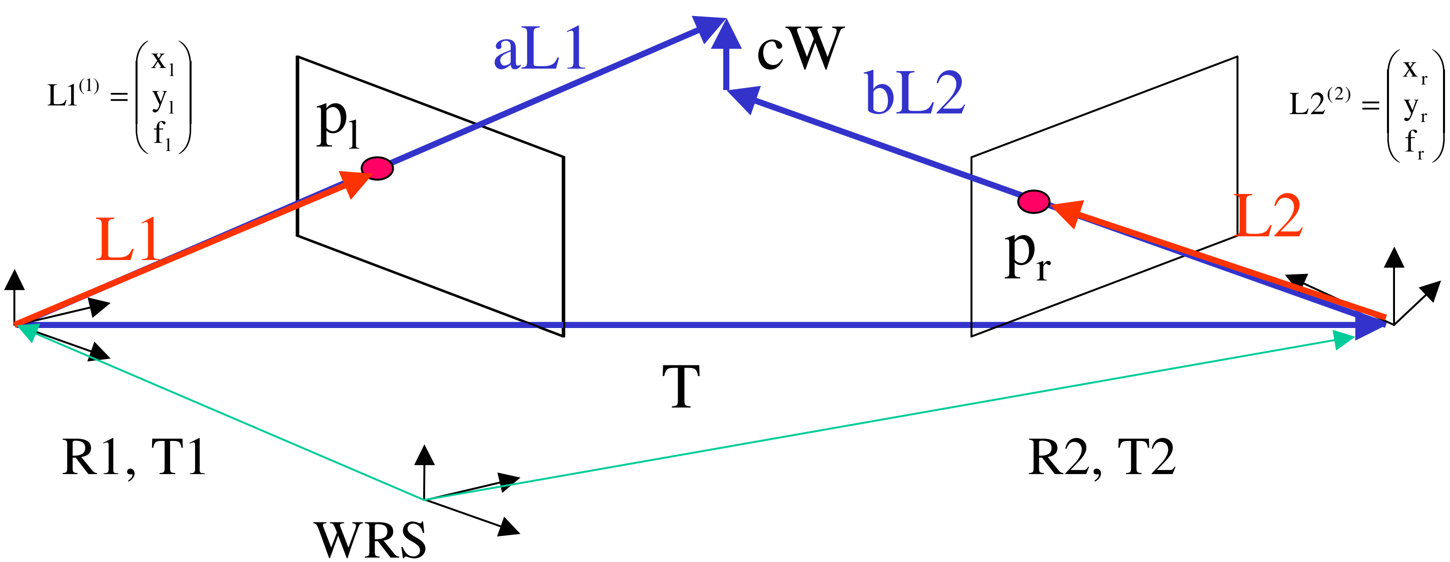
$$\mathbf{w} = \mathbf{L1} \times \mathbf{R}^{-1}\,\mathbf{L2}^{(2)}$$

$$a\,\mathbf{L1} - c\mathbf{W} - \left[(b\,\mathbf{R}^{-1}\,\mathbf{L2}^{(2)}) + \mathbf{T}\right] = \mathbf{T}$$

$$a\,\mathbf{L1} - b\,\mathbf{R}^{-1}\,\mathbf{L2}^{(2)} - c(\mathbf{L1} \times \mathbf{R}^{-1}\,\mathbf{L2}^{(2)}) = \mathbf{T}$$

⇐ Linear system of equations

Relations between the parameters of a stereo system



$$R = R1R2^T$$

$$T = T1 - R1^T T2$$

Reconstruction for intrinsically calibrated cameras (1)

Assuming that:

- the intrinsic parameters are known, and
- n point correspondences are given

the Fundamental Matrix (F) can be computed up to an unknown scaling factor (remember the Eight Point Algorithm results in an homogeneous linear system)n and the Essential Matrix (E) obtained by the equation:

$$F = (M_r^{-1})^T \, E \, M_l^{-1}$$

Where M_l and M_r - matrix for the intrinsic camera parameters

As the Essential matrix is known up to a scaling factor, 3D points can not determined unambiguosly.

Reconstruction for intrinsically calibrated cameras (2)

The matrix E depends on the geometric parameters R and T of the stereo system and is written:

$$E = RS \qquad \text{whit} \qquad S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

$$E\,E^T = [T_x]\,R\,R^T[T_x]^T = \begin{bmatrix} T_y^2 + T_z^2 & -T_x T_y & -T_x T_z \\ -T_x T_y & T_x^2 + T_z^2 & -T_y T_z \\ -T_x T_z & -T_y T_z & T_x^2 + T_y^2 \end{bmatrix} \quad (1) \qquad Tr(E\,E^T) = 2\|T\|^2 \quad (2)$$

so that dividing the entries of the essential matrix by

$$N = \sqrt{Tr(E\,E^T)/2} \quad (3)$$

is equivalent to normalizing the length of translation vector to unit

Reconstruction for intrinsically calibrated cameras (3)

Using this normalization $T_x^2 + T_y^2 + T_z^2 = 1$

$$E\,E^T = [T_x]R\,R^T[T_x]^T = \begin{bmatrix} T_y^2 + T_z^2 & -T_x T_y & -T_x T_z \\ -T_x T_y & T_x^2 + T_z^2 & -T_y T_z \\ -T_x T_z & -T_y T_z & T_x^2 + T_y^2 \end{bmatrix} \quad (1)$$

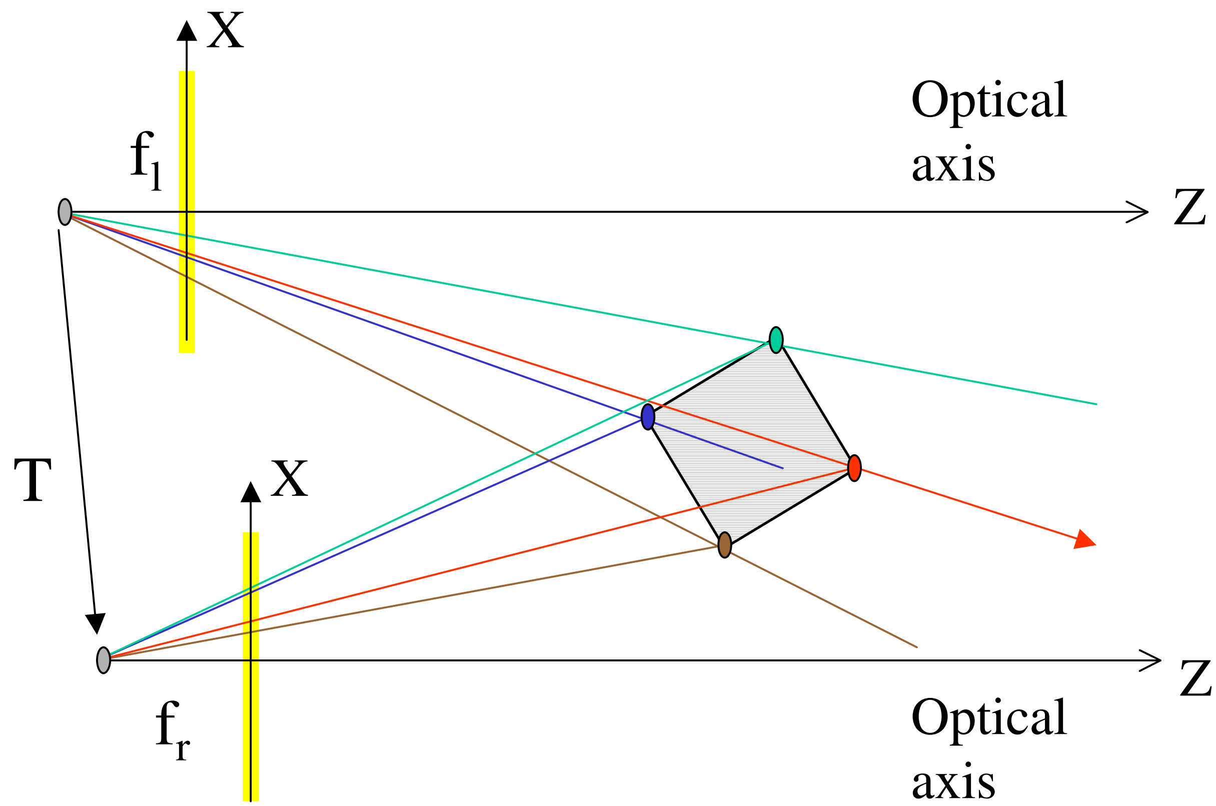
and equation (1) can be rewritten

$$\hat{E}\,\hat{E}^T = \begin{bmatrix} 1 - \hat{T}_x^2 & -\hat{T}_x \hat{T}_y & -\hat{T}_x \hat{T}_z \\ -\hat{T}_x \hat{T}_y & 1 - \hat{T}_y^2 & -\hat{T}_y \hat{T}_z \\ -\hat{T}_x \hat{T}_z & -\hat{T}_y \hat{T}_z & 1 - T_z^2 \end{bmatrix} \quad (3)$$

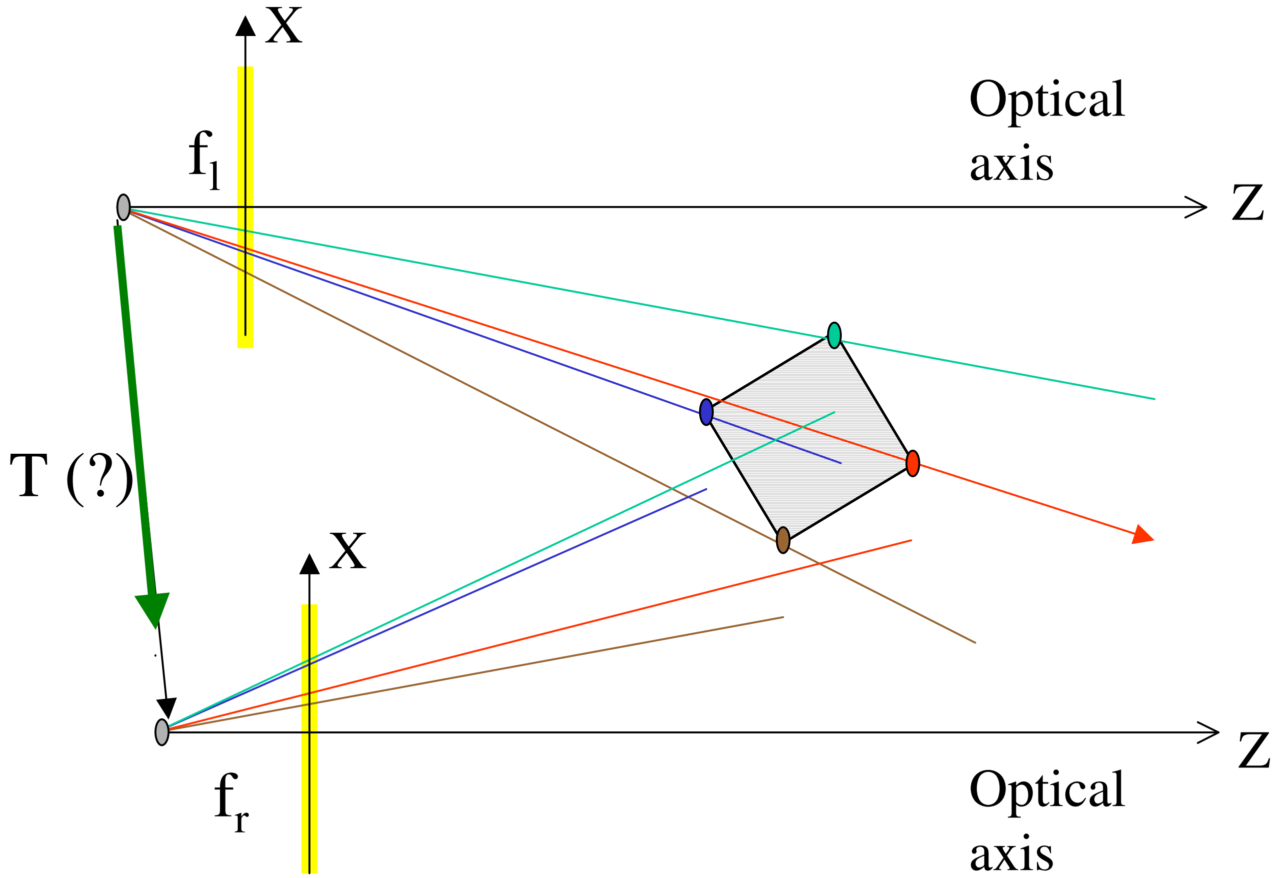
and components of the translation vector (T) can be computed up to sign and rotation matrix determined

For the complete solution see: Introductory techniques for 3D Computer Vision, Emanuele Trucco e Alessandro Verri. Prentice Hall. 1998, section 7.4

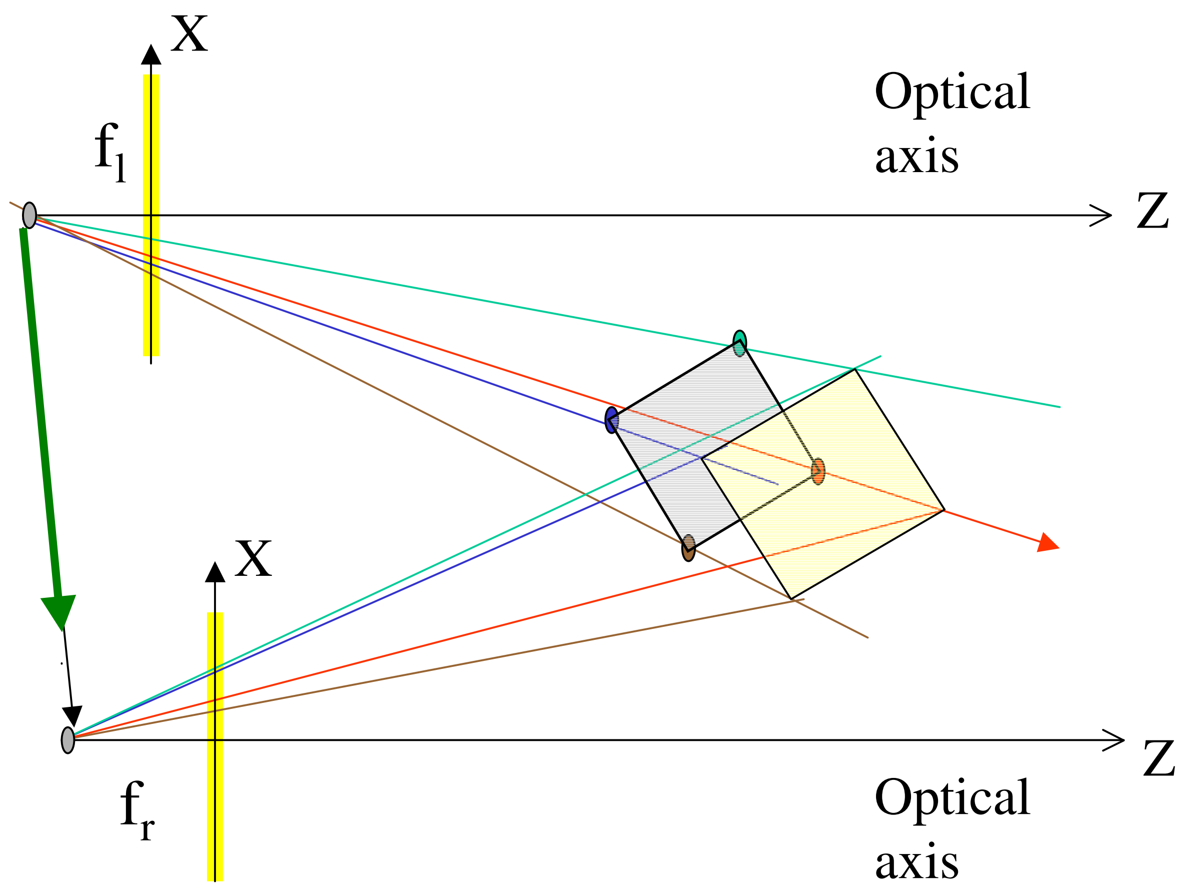
Reconstruction for intrinsically calibrated cameras (4)



Reconstruction for intrinsically calibrated cameras (5)



Reconstruction for intrinsically calibrated cameras (6)



Reconstruction for uncalibrated cameras (1)

Assuming that:

- n point correspondences are given

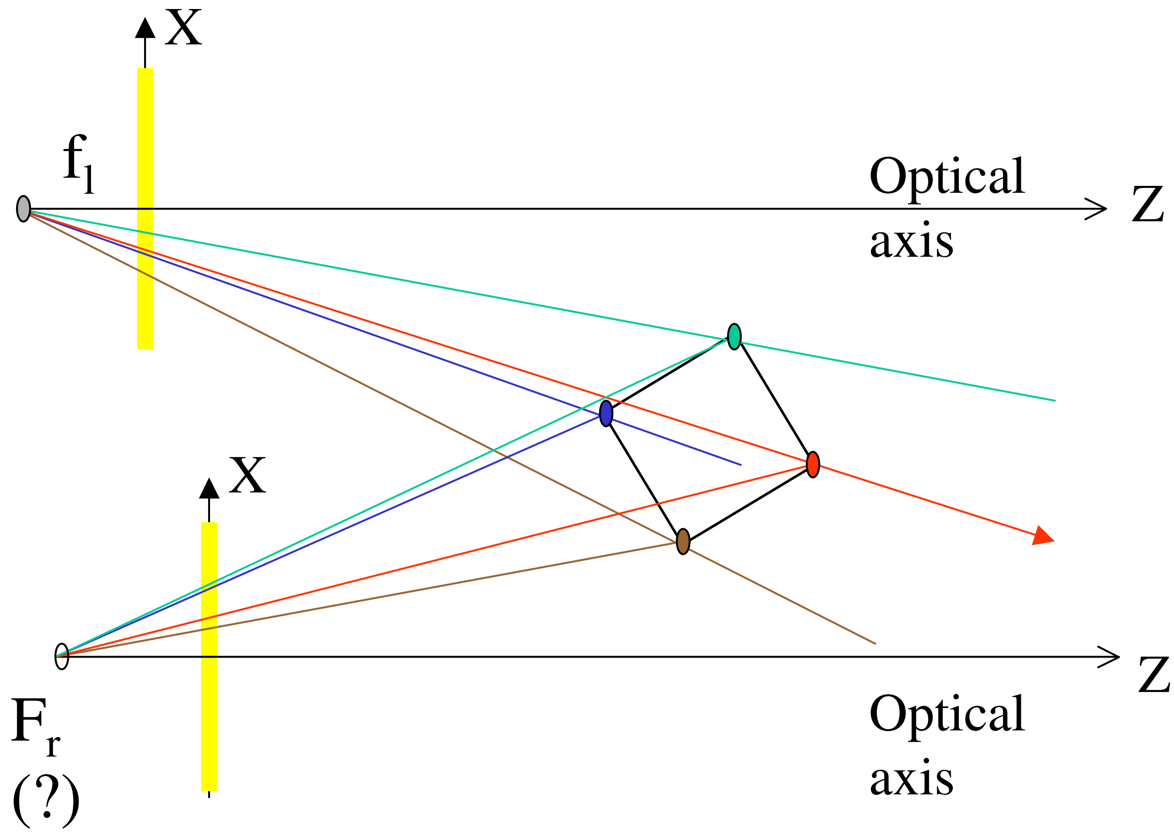
the Fundamental Matrix (F) can be computed up to an unknown scaling factor

As the matrices M_l and M_r for the intrinsic camera parameters are unknown the matrix E can`t recovered from the equation:

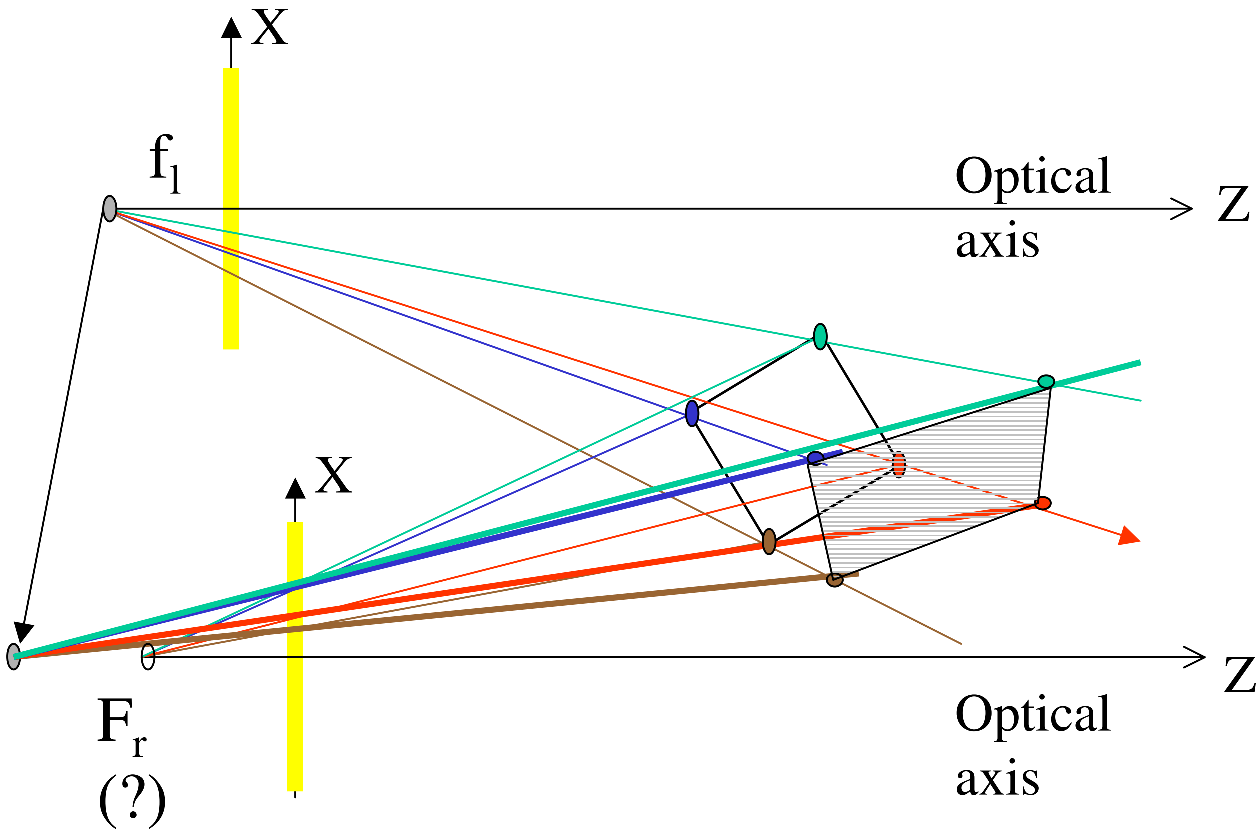
$$F = (M_r^{-1})^T \, E \, M_l^{-1}$$

And the points in 3D space can only be determined up to a projective transformation H

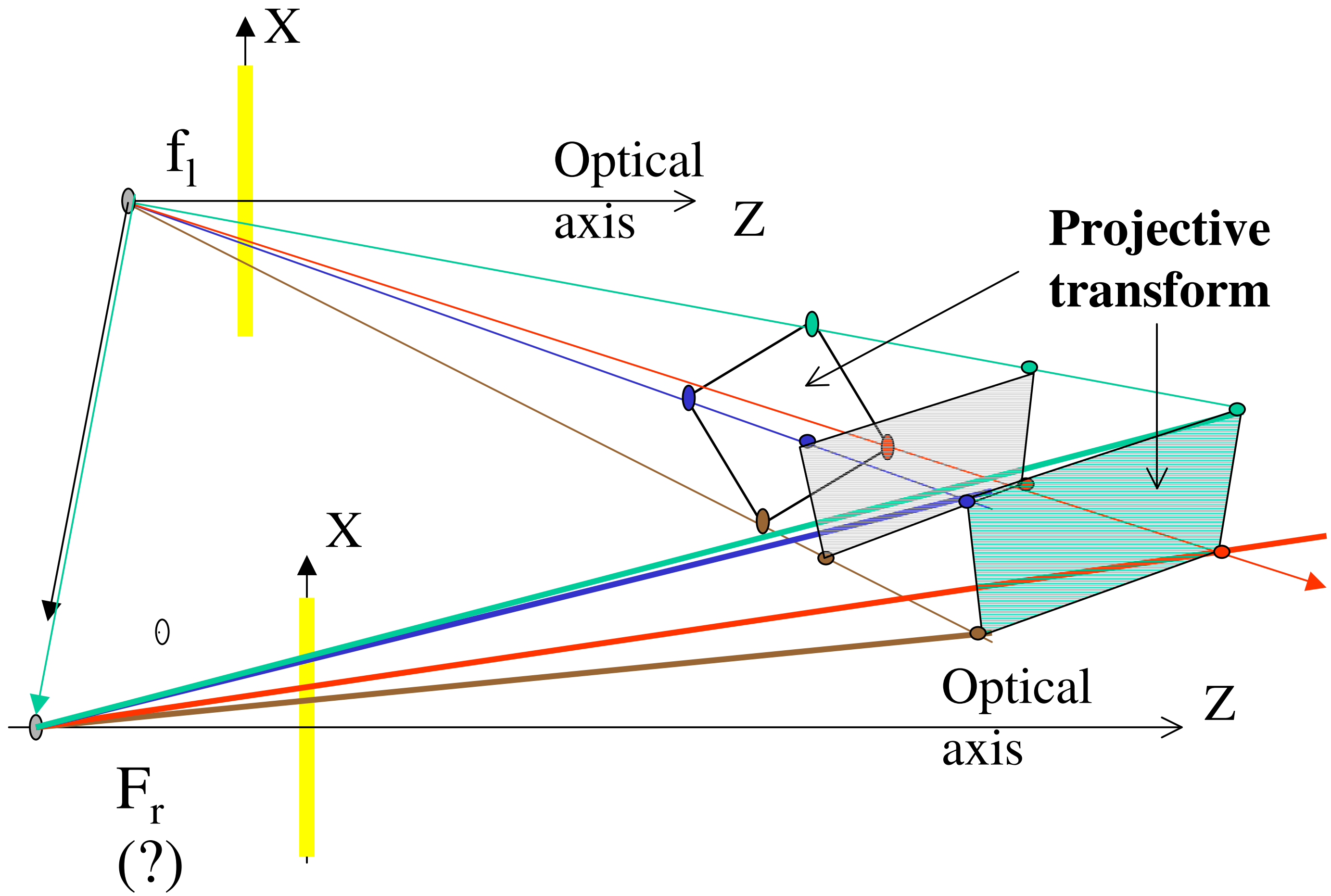
Reconstruction for uncalibrated cameras (2)



Reconstruction for uncalibrated cameras (3)



Reconstruction for uncalibrated cameras (4)

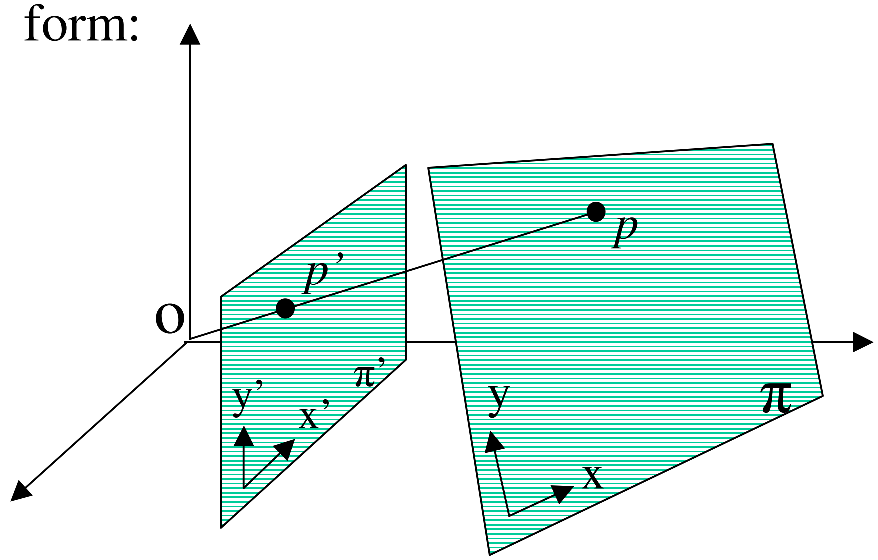


Reconstruction for uncalibrated cameras (5)

Given a point in space in homogeneous coordinate (x,y,z,w) and its image under a projective transform (x',y',z',w'), a projective transform has the following form:

$$\begin{bmatrix} x' \\ y' \\ z' \\ w' \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}^{-1} \begin{bmatrix} x' \\ y' \\ z' \\ w' \end{bmatrix}$$

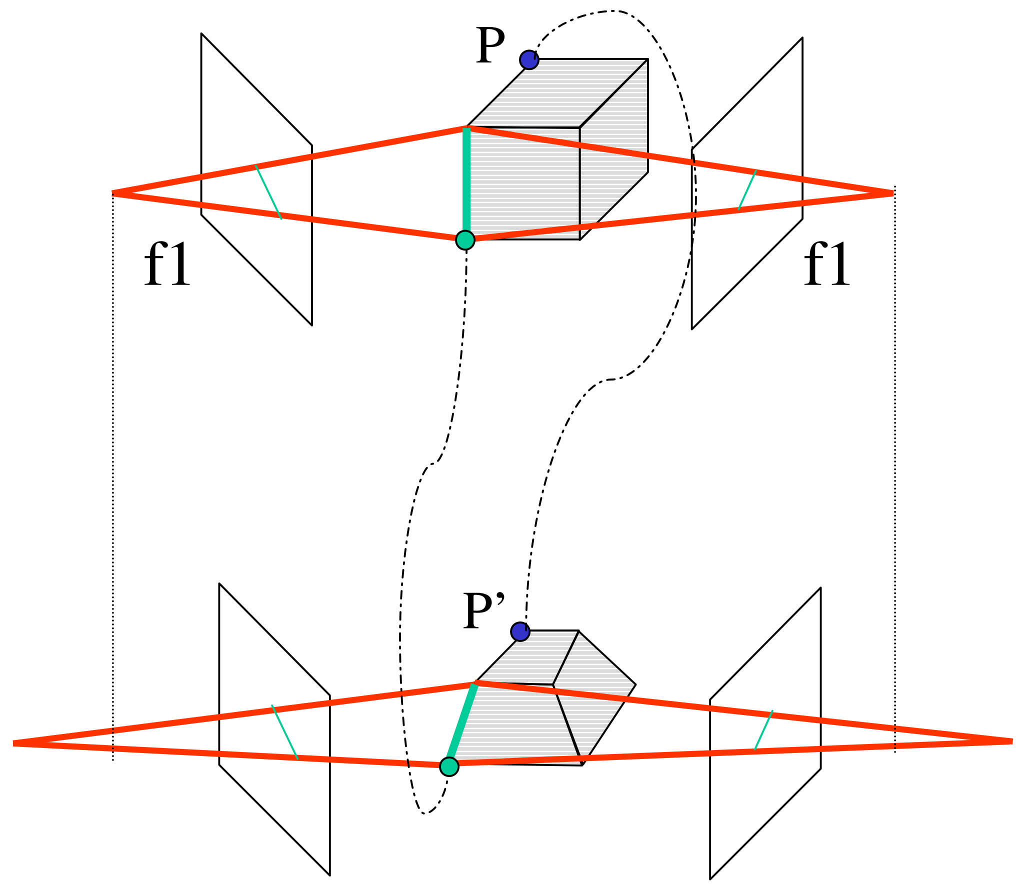


Reconstruction for uncalibrated cameras (6)

Actual object

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \\ w' \end{bmatrix}$$

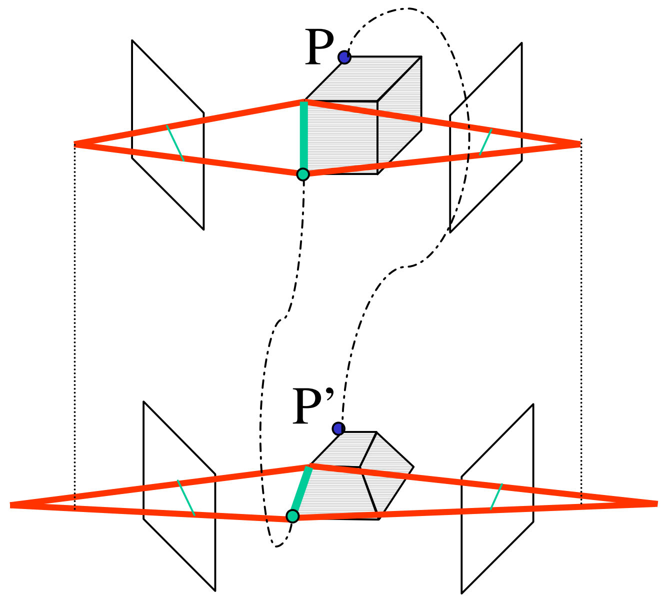
Reconstructed object



Reconstruction for uncalibrated cameras (7)

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \\ w' \end{bmatrix}$$

The projective transformation can be estimated if the coordinates of five 3D points are known and the object reconstructed up to a scaling factor



For the complete solution see: Introductory techniques for 3D Computer Vision, Emanuele Trucco e Alessandro Verri. Prentice Hall. 1998, section 7.4

References

- **Introductory Techniques for 3D Computer Vision**, Emanuele Trucco and Alessandro Verri. Prentice Hall. 1998.
- **Computer Vision**, Linda Shapiro and George Stockman, Prentice Hall, 2001.
- **Computer Vision - A modern approach**, David Forsyth and Jean Ponce , Prentice Hall, 2003.
- **Robot Artificial Vision for Mobile Robots**, Nicholas Ayache. MIT Press, 1991.
- **Vision**, Berthold Klaus and Paul Horn. The MIT Press. 1986.
- **Computer Vision**, Dana H. Ballard and Christopher M. Brown. Prentice Hall. 1982.
- **Digital Image Processing**, Rafael C. Gonzales and Richard E. Woods. Addison Wesley. 1992.

- The Computer Vision Homepage - http://www-2.cs.cmu.edu/~cil/vision.html